UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE CIÊNCIA DA INFORMAÇÃO

Programa de Pós-Graduação em Ciência da Informação

MINERAÇÃO DE TEXTOS E GESTÃO DO CONHECIMENTO: Aplicação na Experiência Operacional em Geração de Energia Nuclear nas Usinas de Angra I e II

ANDRÉ MOREIRA PINTO

BELO HORIZONTE
2007

André Moreira Pinto

MINERAÇÃO DE TEXTOS E GESTÃO DO CONHECIMENTO:

Aplicação na Experiência Operacional em Geração de Energia Nuclear nas Usinas de Angra I e II

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da UFMG, como requisito parcial para a obtenção do titulo de doutor em Ciência da Informação.

Orientação: Professor Doutor Jorge Tadeu de Ramos Neves

BELO HORIZONTE
ESCOLA DE CIÊNCIA DA INFORMAÇÃO DA UFMG
2007

DEDICATÓRIA

À minha esposa, Leda, doce presença luminosa em minha existência, pelo amor, pelo carinho, pela dedicação e paciência e, sobretudo, por ser quem é: pessoa tão especial que escolhi para compartilhar a minha vida.

À minha filha, Beatriz, presente maior do amor de minha esposa e que, tão pequenininha nesse momento de recém-nascida, me faz tão grande e tão feliz como pessoa.

Aos meus pais, pelo amor e carinho sem limites.

AGRADECIMENTOS

Sobretudo a Deus, por me permitir existir e iluminar e guiar a minha vida.

Ao meu orientador, Prof. Dr. Jorge Tadeu de Ramos Neves, pela orientação, incentivo, confiança a mim dedicados e pela competência a ser referenciada.

Aos meus irmãos, Raquel e Marcos e meu sobrinho Rafael, pela paciência.

À Professora Regina Fátima Gibson Gonçalves, pelos ensinamentos valiosos, pela amizade, parceria e carinho fraternal.

Ao Marcelo Gibson de Castro Gonçalves, pela amizade, companheirismo e dedicação no esforço em converter os modelos teóricos de mineração de dados em rotinas computacionais complexas, contribuindo decisivamente para realização dos experimentos realizados nessa Tese.

Ao Sr. Luiz Celso de Oliveira, funcionário da ELETRONUCLEAR, Coordenador da Experiência Operacional, pela receptividade carinhosa, pela competência perspicaz em perceber a potencialidade do conhecimento humano e do trabalho desenvolvido nessa Tese de Doutorado, bem como pelo acolhimento nas várias idas a Angra dos Reis.

Aos funcionários da ELETRONUCLEAR que contribuíram com este trabalho, especialmente os Srs. Carlos Maurício Cruz Belo e Ricardo Lowe Stukart pelo apoio no grupo de pesquisa.

Ao Padre Benjamim Carreira de Oliveira (*in memoriam*) pelo exemplo e amizade inesquecíveis e por ter despertado em mim o gosto e a busca pelo conhecimento.

Ao "cãopanheiro Lui" e à "cãopanheira Luna" que apesar das quatro patas me mostraram nas noites e madrugadas de escrita desse trabalho a lealdade e a fidelidade de uma amizade incondicional.

Aos companheiros e colegas de trabalho da Faculdade Cenecista de Sete Lagoas e do Centro de Inclusão Digital e Social, pelo incentivo e paciência.

À ELETRONUCLEAR, pela abertura das portas ao meu trabalho, pela confiança em mim depositada e pela receptividade às minhas idéias e propostas.

Aos professores, funcionários e colegas da ECI/UFMG que contribuíram com o presente trabalho.

À Universidade Federal de Minas Gerais, pela oportunidade de nela realizar tantas conquistas, descobrir tantos conhecimentos e comprometer-me com a transformação da sociedade brasileira que a mantém e que me proporcionou oportunidades relevantes através dessa Instituição para que eu pudesse nela realizar minha Graduação, meu Mestrado e agora, meu Doutorado.

A Mudança e o Conhecimento

Não tenho dentro de mim um núcleo pequeno e duro, uma estátua completamente formada, real e autêntica, permanente e fixa.

Ser pessoa implica um processo dinâmico. É bom que quanto mais conhecermos a vida, mais podemos mudar. E sempre podemos conhecer mais. E mudar mais ainda.

Por outras palavras, se me conheceu ontem, não vá pensar que é a mesma pessoa que está a encontrar hoje. Experimentei mais da vida, encontrei mais coisas nas pessoas que amo, sofri e supliquei, e estou diferente.

Nem tente atribuir-me um valor médio fixo e irrevogável, porque estou atento o tempo todo, aproveitando cada oportunidade do meu dia-a-dia.

Então quando me encontrar, aproxime-se de mim com um sentido de quem pergunta a meu respeito. Estude o meu rosto, mãos e voz, procurando os sinais de mudança.

Pois é certo que mudei.

(Sabedoria de um autor desconhecido)

RESUMO

O conhecimento é um diferencial importante para que as organizações alcancem seus objetivos. Ele possui uma dimensão tácita e uma dimensão explicita, sendo que a elaboração de documentos e relatórios, artigos e relatos livres são formas acessíveis à maioria das pessoas para tornar explicito o conhecimento que possuem. Nas duas últimas décadas, especialmente após a intensificação do uso de mídias digitais e do uso da Internet as organizações constataram o rápido e elevado crescimento de seus acervos de documentos textuais em formato digital. Em organizações que empregam uso de conhecimento intensivo em suas operações e gestão os documentos textuais possuem uma expressividade enquanto objetos potencialmente portadores de informação e conhecimento ainda maior. Entretanto, extrair e empregar esse conhecimento é um grande desafio, sobretudo quando a organização possui porte elevado e grandes acervos. A descoberta de conhecimento em bases textuais – ou simplesmente mineração de textos – é um relevante recurso para explorar o universo do conhecimento embutido em coleções textuais digitais, permitindo construir visões estratégicas e relevantes a partir do tratamento dessas bases com diferentes recursos computacionais e da Ciência da Informação para apoiar a Gestão do Conhecimento em organizações, especialmente aquelas de conhecimento intensivo. Produzir informação relevante e gerar conhecimento novo e estratégico é o grande desafio da mineração de textos. A presente Tese estuda tais questões a partir de avaliação da realidade organizacional de empresa do setor de energia elétrica de fonte nuclear: a ELETRONUCLEAR, uma vez que esta possui grandes bases de dados textuais relativas ao conhecimento de sua atividade fim – ou seja: a Experiência Operacional. Foi utilizada a ferramenta de mineração de textos ***Difbrain***®, constatando-se sua adequação à geração de conhecimento novo e relevante através da análise sinóptica dos conjuntos de documentos envolvidos. Considerando uma base de documentos textuais digitais a ferramenta realizou leituras inteligentes desses textos, com análise semântica, agrupamentos de textos comuns e inferências avançadas, utilizando para isso recursos metodológicos e computacionais vinculados às áreas de Recuperação da Informação e Inteligência Artificial. Tecnologias inteligentes são necessárias para acelerar a análise, examinando de forma automatizada os documentos e aferindo o que é efetivamente significativo para o usuário. A partir da análise sinóptica de um grupo de documentos foi possível descobrir relações importantes entre eles e que não seriam percebidas sem o uso da descoberta de conhecimento em bases textuais.

Palavras-Chave: Mineração de textos, descoberta do conhecimento em textos, *text mining*, gestão do conhecimento, recuperação da informação, inteligência artificial, gestão da experiência operacional, conhecimento em energia nuclear.

ABSTRACT

The knowledge is important differential for the organizations to achieve its aims. It to have two dimensions: tacit and explicit. The documents, reports, releases, articles and free texts are accessible forms to most of persons for become explicit the yours topics of knowledge. In last two decades, especially after intensive use of digital mídia and Internet technologies, the organizations noticed the quickly and great growth of digital text databases. In the organizations that to use great level of knowledge in the yours operations and management, the text documents to have great value while objects of information and knowledge. However, to extract and to use this knowledge is great match, especially in large databases of documents. The knowledge discovery in text databases – text mining – is one powerful resource to explore the universe of knowledge enable in digital textual collections and can be to build strategic views of reality elements to set of texts. This is achieving with computation resources and approaches of Information Science for support Knowledge Management in organizations, especially those use of intensive knowledge. To produce relevant information and to generate new and strategic knowledge is the great aim of text mining. This thesis studies these questions, to set analysis of organization reality ELETRONUCLEAR, enterprise of power energy of nuclear fonts, that to have large database of digital textual document to concern of activities of generate nuclear energy. ***Difbrain***®, is text mining software that has been use in process text analysis of ELETRONUCLEAR text databases, to doing intelligent reading of documents and extracting relationships, patterns, relevant contents in texts, across semantic analysis, use of Information Retrieval and Artificial Intelligence technologies in synoptic studies about of set of documents.

Key-words: Text mining, knowledge discovery in text, knowledge management, information retrieval, artificial intelligence, operational experience management, nuclear power knowledge.

LISTA DE FIGURAS

LISTA DE ABREVIATURAS E SIGLAS

AIEA – Agência Internacional de Energia Atômica

CNEN - Comissão Nacional de Energia Nuclear

ELETRONUCLEAR – Eletrobrás Termonuclear S.A.

GC – Gestão do Conhecimento

IA – Inteligência Artificial.

INPO (*Institute of Nuclear Power Operators*).

KDD – *Knowledge Discovery in Databases*

KDT – *Knowledge Discovery in Text*

NRC – *United States Nuclear Regulatory Commission*

ORO – Ocorrência de Relato Obrigatório

PLN – Processamento da Linguagem Natural

RDO – Relatórios de Desvio Operacional – RDO

RE – Relatórios de Evento – RE

RES – Relatórios de Evento Significante – RES

SRI – Sistema de Recuperação da Informação

UFMG – Universidade Federal de Minas Gerais

WNA – Word Nuclear Association (Associação Nuclear Mundial)

WANO (*Word Association of Nuclear Operators*)

SUMÁRIO

PARTE I – INTRODUÇÃO

PARTE II – REFERENCIAL TEÓRICO

PARTE III – DESENVOLVIMENTO DA PESQUISA

PARTE IV – FINALIZAÇÃO

PARTE I - INTRODUÇÃO

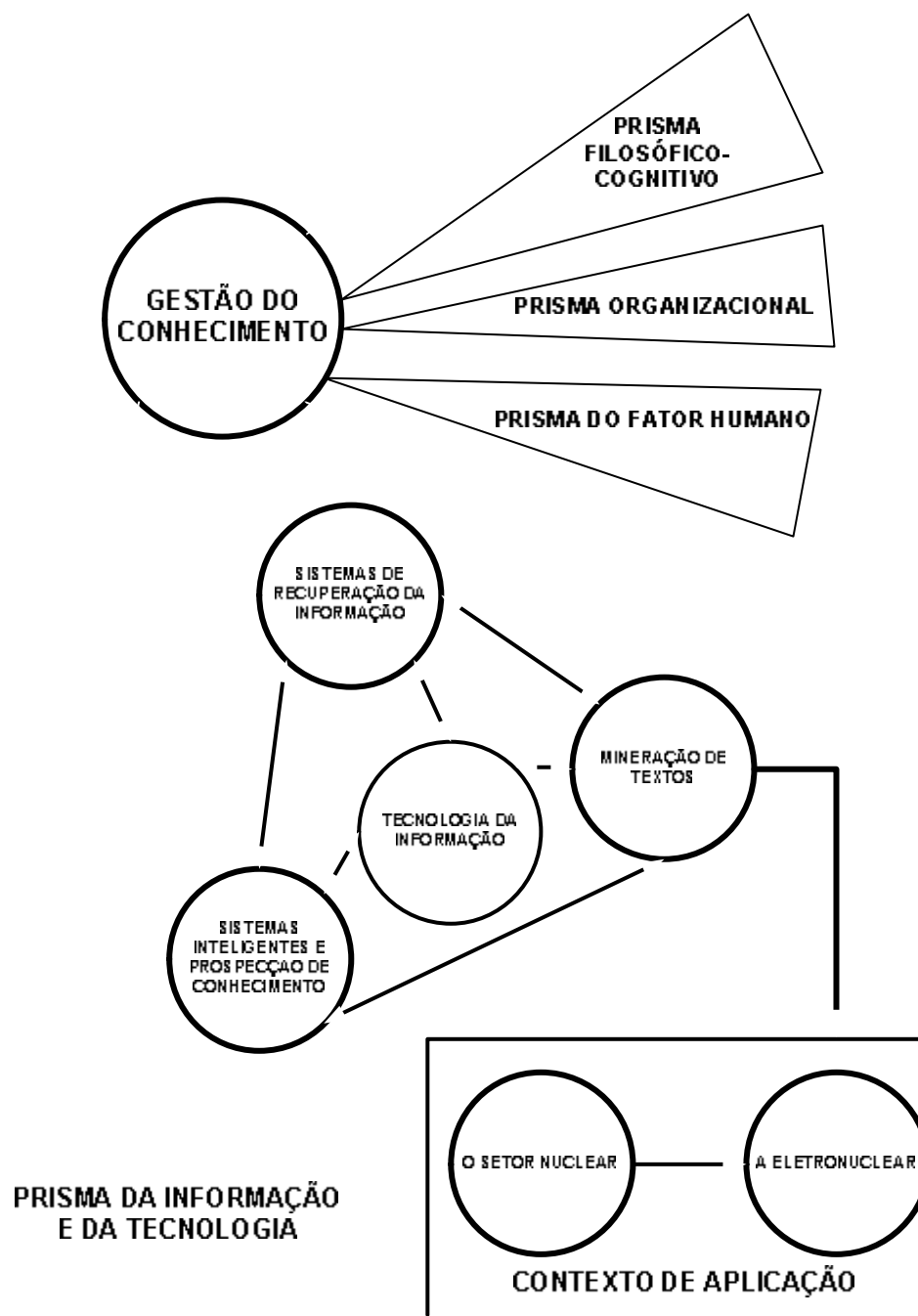
1 CONSIDERAÇÕES INICIAIS

A maior parte das organizações enfrenta hoje uma série de desafios em seu esforço de sobrevivência crescente. A magnitude dos desafios empresariais, sociais, econômicos e políticos colocados pela globalização e uma nova configuração e posicionamento dos agentes econômicos tornam mais complexa a compreensão da realidade e a ação empreendedora sobre ela. Os processos produtivos desenhados numa perspectiva rígida, baseados em atividades mecânicas e repetitivas, em grande parte cederam lugar a novas formas que passaram a demandar volume crescente de informação e conhecimento. Novas capacidades estratégicas e operacionais tornam-se necessárias para as organizações que desejam obter resultados relevantes através da construção de vantagens competitivas e excelência no seu *modus operandi*. A construção dessas vantagens e a busca da excelência passam a ser desafio de toda organização arrojada que busca diferenciação e melhoria contínua no relacionamento com seu público alvo. No caso das empresas, destacam-se aquelas que buscam elementos concretos e eficazes para vencer a concorrência, agora em nível mundial, e conquistar um lugar no mercado. As vantagens competitivas, entretanto, servem a uma causa maior: a sustentabilidade da organização ao longo do tempo. Vantagens competitivas são, em geral, temporárias, mas a necessidade de manter a organização como um empreendimento sustentável é o objetivo maior. Algumas vantagens competitivas, por outro lado, podem ser mais duradouras, embora nem sempre percebidas de forma clara pelos *stakeholders* da organização. Podemos dizer que são vantagens competitivas de raiz, ou seja, sustentam outras vantagens competitivas mais explícitas para clientes e acionistas. É o caso do fator humano e do conhecimento construído por ele na organização.

As vantagens competitivas podem ser construídas através da ação sobre aspectos fundamentais para as organizações, tais como: estratégia empresarial, desenvolvimento e qualidade de produtos, sistemas de produção, sistemas de informações e suas tecnologias agregadas e, em especial, as pessoas. O conhecimento humano sempre foi importante para as organizações, mas atualmente podemos dizer que se tornou a principal vantagem competitiva para grande parte delas. Esse conhecimento é gerado em um contexto no qual se encontram pessoas socialmente articuladas e, cada vez mais, sistemas de informações, os quais, impulsionados pelos avanços tecnológicos, passaram

a apoiar um processo de aceleração da geração e uso do conhecimento, especialmente em melhoria de processos e desenvolvimento de conceitos, inovações e produtos. O binômio pessoas – sistemas de informação resgata e aprofunda uma perspectiva *sócio-técnica* que se mostra pertinente para compreender o desenvolvimento das sociedades complexas contemporâneas, especialmente a partir da segunda metade do século XX. Não obstante o papel do trabalho humano possa estar diminuindo de importância em termos quantitativos, com a crescente automação nas atividades industriais e de serviços, em termos qualitativos para a construção de vantagens competitivas, o mesmo tende a permanecer como sendo muito relevante. As pessoas continuam a ser um elemento central na existência das organizações, sejam as que estão dentro (os empregados), sejam as que estão fora (os clientes) e a percepção correta desse fato pode ser a diferença entre o êxito e o fracasso de um empreendimento. As empresas estão inseridas numa atmosfera de alta mutabilidade. A necessidade de respostas rápidas e criativas por parte das organizações passa a ser um imperativo, o que pode ocorrer não apenas com o uso de novas tecnologias, mas, sobretudo, com pessoas devidamente comprometidas com a geração de tais respostas. Destarte, é relevante, hoje, estudar como a articulação das pessoas com os sistemas de informação, que permite a geração de conhecimento, leva a uma postura estratégica adequada, contribuindo para o maior êxito das organizações nas missões para as quais se estabelecem. Para CHOO (2003) a informação é, em essência, a matéria prima do cotidiano do ser humano, que tem cinco sentidos e deve passar pelo crivo da cognição, da emoção e da situação.

A estrutura do presente trabalho inicia-se com as considerações iniciais descrita no presente capítulo. O capítulo 2 apresenta a justificativa e os objetivos (geral e específicos) do trabalho de tese para sua realização, indicando a relevância do estudo. O capítulo 3 discorre sobre e a metodologia empregada. Os capítulos 4, 5, 6, 7 e 8 apresentam os marcos teóricos do estudo, abordando a gestão do conhecimento, a recuperação da informação, os sistemas inteligentes e a descoberta de conhecimento em bases textuais, respectivamente. O capítulo 9 descreve o contexto da Eletronuclear, caracterizando os desafios setoriais que enfrenta, enquanto o capítulo 10 relaciona os processos de gestão do conhecimento e de uso de tecnologia da informação presentes na empresa. Na seqüência, o capitulo 11 analisa a mineração de textos na Eletronuclear e são apresentadas posteriormente, no capítulo 12 as considerações finais com uma conclusão e perspectiva futura. Na seqüência são apresentadas as referências e os

anexos. A Figura 1 a seguir apresenta uma visão geral da proposta dessa tese. O foco principal é a questão do conhecimento humano, considerada e estudada por quatro prismas diferentes, filosófico-cognitivo, organizacional, fator humano e da informação e tecnologia com destaque para este último, em que a mineração de textos é aplicada num contexto organizacional específico: o da ELETRONUCLEAR.

2 MOTIVAÇÃO E OBJETIVOS

2.1 Justificativa

2.1.1 Aspectos Preliminares

O tema gestão do conhecimento, nos últimos anos, está absorvendo a atenção de muitas organizações, especialmente daquelas que possuem visão estratégica, estruturas amplas e complexas na geração e comercialização de bens e serviços e, bem como das buscam continuamente aperfeiçoar as respostas e resultados oferecidos aos seus clientes e comunidades nas quais se inserem. Iremos assumir, para efeito desse trabalho, o importante vínculo existente entre conhecimento e informação, especialmente a informação textual, dentro das organizações. Conhecimento e informação expressam realidades diferentes que serão exploradas ao longo desse trabalho em uma perspectiva não exaustiva. A importância do presente trabalho vincula-se especialmente à geração automática de informação a partir de bases textuais, visando apoiar a gestão do conhecimento. Para indicar a importância dessa perspectiva, será desenvolvida uma pesquisa-ação vinculada a um conjunto de experimentos com uso de software em um contexto complexo, expresso pela atividade empresarial da ELETRONUCLEAR.

2.1.2 – Problemas e Desafios da Informação Textual

Na maior parte das organizações a quantidade de documentos textuais é grande. Muitas consultas a esses documentos são necessárias na dinâmica do dia-a-dia de muitas empresas e a busca por tais documentos pode ser muitas vezes lenta devido à diversidade de temas tratados por tais documentos e à quantidade deles que está armazenada, muitas vezes em múltiplos locais. Além disso, muitas vezes documentos armazenados em locais diferentes podem estar relacionados uns com os outros e é preciso identificar tal relação para solução de muitas questões empresariais. O tempo é sempre um fator importante. Muitas vezes, em uma pesquisa, localiza-se um grande conjunto de documentos de interesse, mas apenas uma pequena parte é realmente relevante. Relevância é uma questão importante para a recuperação de informações. Um sistema de recuperação da informação só pode apresentar informações relevantes se conseguir identificar o que significa ser relevante para o usuário. E isso não é simples.

O que é relevante para uma pessoa pode não ser para outra. Informação relevante é aquela informação que o usuário necessita em determinado momento para a realização de alguma coisa, ou seja, ela deve estar no contexto que o usuário quer e no momento certo. Para SARACEVIC (1970) o conhecimento é adequadamente comunicado quando há mudanças e a relevância é a medida dessas mudanças.

Em um passado recente, na maior parte das organizações, os dados textuais, embora em volume significativo, quase não eram considerados como elemento de sustentação para geração de informação gerencial e definição de vantagem competitiva. De igual forma, embora intuitivamente fosse percebida a importância da informação textual para gestão do conhecimento organizacional, o manejo de grandes volumes de textos dificultava o uso da mesma em termos práticos. Os editores de texto e o ambiente *web*, bem como os processos que vem sendo denominado por mineração de textos (*Text Mining*) permitiram mudar essa perspectiva. Se por um lado, editores de texto e páginas *web* garantem facilidades ao usuário na geração de informação e expressão do conhecimento, as ferramentas de mineração viabilizam inferências inteligentes sobre texto e sua progressiva exploração como recurso informacional cada vez mais relevante para as organizações. Tecnologias se apresentam como necessárias para acelerar a análise dos textos, examinando de forma automatizada os documentos e aferindo aquilo que é verdadeiramente significativo. É possível também, a partir da análise sinóptica de um grupo de documentos, descobrir relações importantes entre eles e que antes não seriam percebidas.

2.2 Problema de Pesquisa e Pressupostos

2.2.1 Problema de Pesquisa

Considerando a necessidade de se “extrair conhecimento” de grandes acervos em formato digital compostos por bases de documentos textuais neles armazenados, no seio de uma organização, o problema de pesquisa que a tese pretende tratar é o de como utilizar a estratégia tecnologia da descoberta de conhecimento em textos – mineração de textos – como suporte diferenciado ao processo de gestão do conhecimento organizacional. Para tanto, coloca-se o imperativo da viabilização e utilização inteligente e tempestiva de coletâneas de textos digitalizados na geração de

conhecimento relevante para essa organização, através de ferramenta de software que atue com referenciais semânticos e estatísticos os quais incluam o uso de estratégias de recuperação da informação, recursos de inteligência artificial e uso de elementos computacionais, todos integrados sob o prisma de um projeto de gestão do conhecimento.

Considera-se que a elaboração de textos, livres (artigos, descrições, mensagens ou correspondências) ou semi-estruturados (sob a forma de relatórios e formulários,) é uma forma de converter parcelas importantes de conhecimento tácito em conhecimento explicito, dada a relativa facilidade de sua elaboração por parte de múltiplos usuários que podem fazê-lo com elevado grau de liberdade, dentro de estilos, intenções e visões de forma como melhor entendem. O processamento de linguagem natural, processo de inteligência artificial embutido em softwares de mineração de textos, contribui para o tratamento de coleções textuais digitais, permitindo o alcançar a descoberta de significados (aspecto semântico) com a análise de textos.

2.2.1 Pressupostos

Os pressupostos vinculados à abordagem do problema são:

- Os textos são expressões do pensamento humano com grande flexibilidade e grau de liberdade para elaboração por parte das pessoas, tornando-se assim um modo acessível de conversão de conhecimento tácito para conhecimento explicito em muitos contextos organizacionais.

- A mineração de textos dá às coleções textuais digitais armazenadas nas organizações um novo e significativo potencial de geração da informação e conhecimento.

- É possível utilizar ferramenta informatizada, baseada em abordagens semânticas e estatísticas para responder de forma inovadora e eficaz as necessidades de geração de informação útil para o usuário.

- As informações geradas pela ferramenta de mineração de textos, por sua capacidade de análise semântica e possibilidade de construção de visões sinópticas de grupos de documentos (derivadas especialmente dos processos de sumarização, categorização e *clustering*) são relevantes para a geração de conhecimento novo para o usuário, satisfazendo-o em suas necessidades.

- A mineração de textos, usando recursos de inteligência artificial, é uma tecnologia que apresenta potencialidades e ganhos visíveis para o usuário, podendo cumprir de maneira inovadora o suporte à gestão do conhecimento em ambientes complexos que incluam a necessidade de manipular grandes volumes de informação textual.

2.3 Objetivos

2.3.1 Objetivo Geral

Apontar a contribuição da tecnologia de mineração de textos – *text mining* - para o aprimoramento da gestão do conhecimento, propondo, implantando e avaliando os resultados de modelo de recuperação automática da informação textual digital em organizações fortemente dependentes de conhecimento, inovação bem como de informações estratégicas e operacionais para tomada de decisão e excelência na atuação.

2.3.2 Objetivos Específicos

- Desenvolver os aspectos conceituais do modelo de mineração de textos proposto, descrevendo seus fundamentos, objetivos e características e indicando a relevância dessa tecnologia como suporte à gestão do conhecimento.

- Definir e descrever os recursos de mineração de textos que deverão ser utilizados em um sistema informatizado, apontando a relevância do desenvolvimento de software adequado a tal função.

- Atuar no desenvolvimento de sistema de mineração de textos, junto a grupo de trabalho especializado, participando de seu projeto, desenvolvimento e teste e

viabilizando a recuperação automática de informação textual para detecção inteligente de conhecimento embutido em bibliotecas digitais.

- Entender e caracterizar as necessidades de informação de uma organização alvo de pesquisa, que seja fortemente dependente de conhecimento e inovação e de informações estratégicas e operacionais para tomada de decisão, especialmente no que tange a questões gerenciais relevantes sobre as quais a mineração de textos pode redundar em benefícios importantes, consistentes e robustos.

- Conduzir projeto de experimentos que permita avaliar e monitorar a utilização do sistema de mineração de textos na organização pesquisada, avaliando seus resultados e explicitando de que forma tal tecnologia pode contribuir para a gestão do conhecimento e o desenvolvimento de uma postura estratégica na mesma.

3 METODOLOGIA

3.1 Visão Geral

Conforme THIOLLENT (1986):

> "*a metodologia é entendida como disciplina que se relaciona com epistemologia ou a filosofia da ciência. Seu objetivo consiste em analisar as características dos vários métodos disponíveis, avaliar suas capacidades, potencialidades, limitações ou distorções e criticar os pressupostos ou as implicações de sua utilização. Ao nível mais aplicado, a metodologia lida com a avaliação das técnicas de pesquisa e com a geração de experimentação de novos métodos que remetem aos modos efetivos de captar e processar informações e resolver diversas categorias de problemas teóricos e práticas de investigação. Além de ser uma disciplina que estuda os métodos, a metodologia é também considerada um modo de conduzir a pesquisa* (p.25)"

Em geral, numa visão clássica, o processo de uma pesquisa envolverá a definição de uma questão ou problema de pesquisa que depois será traduzida em hipóteses ou pressupostos, os quais serão, por sua vez, verificados. A definição do problema envolverá conhecimentos e valores do pesquisador, ou seja, há dimensões teóricas e éticas compondo seu quadro pessoal de referência. Uma vez consciente do problema, ou questão da pesquisa, o pesquisador procurará delimita-lo elaborando hipóteses, para as quais em geral revisará a produção intelectual já existente (revisão bibliográfica e busca de outros dados relevantes em artigos, jornais, revistas, Internet, etc.). De acordo com a questão de pesquisa definida e com as hipóteses formuladas (vale dizer também, de acordo com a consistência inicial dessas hipóteses) o pesquisador adotará determinadas estratégias de verificação. Hipóteses diversas implicam em diferentes necessidades

De acordo com BERTO & NAKANO (1999) o primeiro passo para se projetar uma boa pesquisa é determinar qual o seu propósito, que pode ser exploratório, descritivo, explanatório ou preditivo. A determinação das estratégias de verificação levará isso em consideração, redundando assim nesse ou naquele método de pesquisa bem como num ou noutro instrumento de coleta de dados. Essas estratégias de verificação são baseadas em métodos de abordagem quantitativa e/ou qualitativa, embora muitas pesquisas possam mesclar algumas posturas. Ainda de acordo com BERTO & NAKANO (1999):

> *“a metodologia de pesquisa provê subsídios ao planejamento e desenvolvimento sistematizado de uma investigação científica a respeito de um fenômeno observado na “realidade do mundo físico material”. Utiliza um ou vários métodos combinados de observação, de maneira a apreender fatos e dados dessa realidade, com a intenção de entender, explicar e, se possível ou necessário, aplicá-la ou replicá-la em favor de outros eventos ou episódios semelhantes. Inicialmente, a dificuldade de escolha ou opção por um determinado método está ligado à dificuldade em se estabelecer ou identificar claramente o problema (ou solução) que se quer verificar ou conhecer. A pesquisa bibliográfica permite construir (ou reconstruir) um modelo implícito nos diferentes constructos, criando um arcabouço teórico capaz de sustentar ou subsidiar as questões de pesquisa* (p.2)*”.*

As perspectivas modernas de pesquisa repousam sobre duas grandes vertentes conceituais: o racionalismo e o empirismo. Na primeira, segundo VARGAS (1985), a visão de Descartes assumiu que através da razão, do raciocínio lógico-dedutivo é possível chegar-se à verdade sobre um fato, baseando-se nos princípios de igualdade entre verdade e evidência e da divisão de um problema em pequenas partes para ser analisado. Já na perspectiva empirista a importância recai sobre a observação da realidade, com uso de um raciocínio indutivo, através do qual pela observação dos fatos, sem preconceitos, é possível se chegar a uma lei geral. Modernamente, de acordo com BERTO & NAKANO (1999) a ciência vem se servindo de uma conduta de pesquisa que consiste de uma mesclagem das duas perspectivas e que dá origem ao método empírico, ou pesquisa quantitativa. Segundo BRYMAN (1989: 6-9), essa conduta enfatiza a formulação de hipóteses sólidas, a partir de um marco teórico, da qual se extraem conceitos passíveis de serem medidos e verificados (operacionalização). Ainda segundo esse autor, busca-se com as hipóteses a demonstração de uma relação de causa e efeito, seja de forma explícita ou implícita e a pesquisa deve-se preocupar com a construção de conclusões que possam ser generalizadas e novamente verificadas por outro pesquisador que utilize os mesmos pressupostos e procedimentos.

Uma pesquisa quantitativa em geral se apresenta adequada quando é possível realizar medidas quantificáveis de variáveis e inferências a partir de amostras de uma população. Esse tipo de pesquisa usa medidas numéricas para testar visões científicas, hipóteses e proposições. O Projeto de Experimentos e a realização de Pesquisa de *Survey* são exemplos de métodos de pesquisas quantitativas.

Uma outra conduta de pesquisa é dada pela abordagem qualitativa ou interpretativa, que é caracterizada por um nível baixo ou inexistente de medidas numéricas ou análises estatísticas. Nesse tipo de pesquisa, examinam-se questões e aspectos mais profundos, com consideração de elementos subjetivos de um objeto de estudo.

CALDER (1977) classifica os métodos qualitativos de acordo com seus diferentes possíveis usos, agrupando-os em três abordagens: exploratória, fenomenológica e clínica. A primeira abordagem, a exploratória, pode ocorrer quando um pesquisador se interessa em testar aspectos operacionais de uma pesquisa quantitativa ou quando pretende incentivar e incrementar o pensamento científico através de uma formulação mais profunda de uma determinada questão ou tema.

Uma abordagem fenomenológica é o segundo tipo de possibilidade de uma pesquisa qualitativa. De acordo com BERTO & NAKANO (1999) na pesquisa qualitativa:

> *"o pesquisador procura reduzir a distância entre a teoria e os dados, entre o contexto e a ação, usando a lógica da análise fenomenológica,isto é, da compreensão dos fenômenos pela sua descrição e interpretação".*

De acordo com AAKER (1990) a abordagem fenomenológica, chamada por ele como sendo "de orientação" tem como propósito

> *"transferir o pesquisador para o ambiente que lhe é pouco ou nada familiar, fazendo com que experimente as mesmas sensações, problemas, necessidades e satisfações da população pesquisada. Ao se "transformar" em membro dessa população, o pesquisador aprende sua linguagem, seu vocabulário e adquire subsídios mais consistentes para sua pesquisa, a partir de uma interação muito mais próxima com a realidade, sob o ponto de vista do universo pesquisado* (p.162)".

A fenomenologia está ligada, sobretudo, à produção intelectual de HUSSERL[1]. Segundo HUSSERL (1913) a fenomenologia

---

[1] Edmund Husserl. Filósofo alemão (1859-1938) considerado o teórico fundador da fenomenologia que, de certa forma, retoma modernamente o objetivismo de Platão, no qual as idéias são realidades objetivamente dadas que formam uma ordem factual. Mas HUSSERL se afasta de Platão na medida em que, no lugar de uma concepção mitológica, que pressupõe a existência da alma, põe uma intuição das

> *"é o estudo dos fatos vividos da consciência na sua pura generalidade essencial e não como fatos realmente experimentados e apreendidos empiricamente por seres conscientes".*

A fenomenologia buscará respostas para uma questão crítica do conhecimento: como a nossa consciência pode ter acesso aos objetos? Como se forma, para nós o campo da nossa experiência? HUSSERL irá recusar a noção clássica de conhecimento como representação e partirá da idéia de que o conhecimento surge do encontro entre a subjetividade e os objetos a partir de uma atitude intencional. Isso significa que a consciência é a consciência de alguma coisa e não da representação (imagem ou símbolo) dessa coisa. A consciência apresenta os objetos para o indivíduo e não os representa. ABRÃO (1999: 439-440), analisando a obra de HUSSERL afirmará:

> *"se nossa consciência é intencional, se ela é sempre direção a um objeto, a tarefa primeira da fenomenologia será verificar quais as sínteses que se encontram na origem desse efeito espetacular: temos consciência de um mundo, a despeito da variedade das perspectivas que possamos ter sobre ele. Enquanto em nossa experiência efetiva existe um fluxo permanente de manifestações e fenômenos de objetos, a capacidade sintética de nossa consciência faz com que, por meio dessa multiplicidade de fenômenos, tenhamos consciência de um objeto uno e idêntico"*

Já MERLEAU-PONTY (1945) caracteriza a fenomenologia como:

> *"uma filosofia para a qual o mundo está sempre e já em aí, comantecedência à reflexão, como presença inalienável e cujo esforço é ode recobrar esse contato ingênuo com o mundo para lhe conceder,finalmente, um estatuto filosófico* (Prólogo)*"*

Ao procurar compreender, descrever e interpretar os fenômenos, a fenomenologia irá fornecer suporte teórico para abordagens qualitativas de pesquisa. Para VAN MAANEN (1979) as experiências pessoais do pesquisador são importantes elementos na análise e compreensão dos fenômenos estudados. Para BRYMAN (1989), na pesquisa qualitativa não há hipóteses fortes no início da pesquisa e o pesquisador observa os fatos como alguém interno à organização, sendo que a pesquisa busca compreender melhor o contexto da situação. Para esse autor, a pesquisa deve enfatizar o processo dos acontecimentos e a seqüência de fatos ao longo do tempo.

---

essências condicionada pelo fenômeno concreto e consumada pelo próprio fenômeno. (Cf. HESSEN, 1999: 70-71).

Ainda de acordo com a visão de CALDER (1977), temos também a terceira abordagem de uma pesquisa qualitativa que é a clínica, em que se pretende trazer à tona sensações e sentimentos que não seriam facilmente percebidos por meio de métodos de pesquisa muito estruturados. Lida com informações latentes, por vezes inconscientes e inacessíveis nos relacionamentos inter-pessoais, constituindo-se em uma boa ferramenta na obtenção de informações úteis para o julgamento clínico de um profissional especializado.

Dentre os métodos de abordagem qualitativa podemos citar o Estudo de Caso e a Pesquisa-Ação. Esses métodos utilizam diferentes estratégias de verificação, com instrumentos que podem ser utilizados isoladamente ou combinados entre si. Dentre estes podemos citar as entrevistas (estruturadas, semi-estruturadas ou não-direcionadas), uso de questionários, observação direta, participação direta, técnicas projetivas[2] e os grupos focais.

A presente tese adotou como método principal a Pesquisa-Ação. Porém, houve também sua interseção com recursos do Projeto de Experimentos para desenvolvimento de teste de *software* de mineração de textos. Duas técnicas foram associadas a esses métodos: o Grupo Focal e o Processo Interativo Orientado a Objetos para Desenvolvimento de *Software*.

3.2 A Pesquisa – Ação

A presente Tese utilizou o método da Pesquisa-Ação conforme descrito em THIOLLENT (1986) e que procuraremos descrever resumidamente com base nesse autor. Essa abordagem lida com a ação interativa de um pesquisado dentro de um grupo implicado em uma situação problema, sendo que um dos objetivos desse tipo de

---

[2] Técnicas projetivas são usadas quando o pesquisador considera praticamente impossível aos entrevistados responderem sobre as reais razões que os levam a assumir certas atitudes e comportamentos. Muitas vezes as próprias pessoas desconhecem seus sentimentos e opiniões a respeito de determinado assunto, sendo incapazes de verbalizar seus verdadeiros sentimentos e motivações. O pesquisador, nesses casos, apresenta um estímulo ambíguo aos entrevistados e observa suas reações no relacionamento com esse estímulo. Esse tipo de técnica é bastante utilizado em sessões de psicoterapia, quando o paciente é solicitado a se projetar como se fosse outra pessoa ou objeto e a expressar seus sentimentos sob essa nova identidade .

proposta metodológica é conferir aos pesquisadores e grupos participantes os meios de se tornarem capazes de lidar com situações-problema sob a forma de uma ação transformadora. De acordo com LIEBSCHER (1998), para saber utilizar métodos qualitativos é preciso aprender a observar, registrar e analisar interações reais entre pessoas, e entre pessoas e sistemas. Isso tem muito em comum com a Pesquisa-Ação, método de pesquisa de caráter interativo e fenomenológico e que busca a construção de soluções para certos problemas, com grande envolvimento do pesquisador com a realidade pesquisada. De acordo com THIOLLENT (1986), a pesquisa-ação destina-se a estudos do que chama de "faixa intermediária", ou seja: o estudo de situações que não se encontram nem no nível micro-social (indivíduos e pequenos grupos) e nem no nível macro-social (a sociedade como um todo). Tal faixa intermediária corresponde a uma grande diversidade de atividades de grupos e indivíduos no seio ou margem de instituições ou coletividades, sendo um instrumento de trabalho e de investigação com grupos, instituições e coletividades de pequeno e médio porte, o que é o caso da grande maioria das organizações públicas e privadas. O autor afirma ainda que:

> "*Pesquisa - ação é um tipo de pesquisa social com base empírica que é concebida e realizada em estreita associação com uma ação ou com a resolução de um problema coletivo e no qual os pesquisadores e os participantes representativos da situação ou do problema estão envolvidos de modo cooperativo ou participativo (..)Para que não haja ambigüidade, uma pesquisa pode ser qualificada de pesquisa-ação quando houver realmente uma ação por parte das pessoas ou grupos implicados no problema sob observação. Além disso, é preciso que ação seja uma ação não trivial, o que quer dizer uma ação problemática, merecendo investigação para ser elaborada e conduzida. (...) Na pesquisa-ação os pesquisadores desempenham um papel ativo no equacionamento dos problemas encontrados, no acompanhamento e na avaliação das ações desencadeadas em função dos problemas* ." (p:14-15)

Os principais aspectos dessa abordagem são:

- Interação entre pesquisadores e pessoas envolvidas na situação a ser investigada;
- Prioridades dos problemas como resultado da interação;
- Relação do objeto de investigação com a situação e os problemas encontrados;
- Objetivo da pesquisa-ação: resolução do problema ou seu esclarecimento;
- Há o acompanhamento das decisões, ações e de toda atividade intencional;

- Ampliação do conhecimento de todos inseridos na situação "nível de consciência".

Prosseguindo na descrição apresentada por THIOLLENT (1986), cumpre salientar que no caso da pesquisa-ação em geral não possuímos hipóteses prévias totalmente desenhadas e consolidadas e associadas a um pequeno grupo de variáveis precisas, isoláveis e quantificáveis, como no método científico tradicional. Na pesquisa-ação temos interações entre grupos e dentro de grupos, situações nas quais se manifestam muitas variáveis imprecisas dentro de certo contexto em mudança permanente. Por isso a pesquisa ação opera a partir de certas premissas, instruções ou diretrizes concernentes ao modo de encarar os problemas identificados na situação investigada e relativa aos modos de ação. A partir dos resultados da pesquisa, tais premissas podem sair fortalecidas ou, ao contrário, serem abandonadas ou reformuladas. Tais premissas não obrigatoriamente precisam ser submetidas a experimentos ou testes estatísticos. Sempre que possível podemos utilizar experimentos, mas eles podem ser enfocados em uma ótica também qualitativa e não somente quantitativa ou sujeitos somente à mensuração estatística. Dentro da abordagem de Pesquisa-Ação, a hipótese ou diretriz, ou ainda os pressupostos precisam de uma formulação de modo claro e conciso, sem ambigüidade gramatical e designar os objetos em questão a respeito dos quais seja possível fornecer provas concretas e argumentos convincentes, favoráveis ou não. Para fins descritivos, a hipótese qualitativa, ou diretriz deve ser usada para organização da pesquisa em trono de possíveis conexões ou implicações não causais, mas com precisão suficiente para se estabelecer que X tem a ver com Y na situação considerada (THIOLLENT, 1986, p. 56).

Merece também avaliação a questão das inferências e da generalização na Pesquisa-Ação, ou seja: a passagem de um nível local para o global, conforme THIOLLENT (1986). Conforme o autor:

> "*a inferência é considerada como passo de raciocínio possuindo qualidades lógicas e de controle. A concepção estatística das inferências não esgota toda a complexidade qualitativa das inferências no contexto particular da pesquisa social. As inferências constituem passos do raciocínio na direção da generalização. Isto corresponde à indução. Antes de serem problema de estatística, as inferências são tema de lógica. O seu controle remete ao conhecimento de algumas regras de lógica elementar. (...) Em situações de*

> *pesquisa, os pesquisadores devem estar atentos em não confundir as inferências efetuadas por eles e as inferências efetuadas por outros participantes. Os pesquisadores devem identificar as generalizações populares e cotejá-las com generalizações teóricas. (...) Independentemente das exigências estatísticas e lógicas que podem ser aplicadas nos casos de uma quantificação ou de uma formalização do conhecimento, os pesquisadores aplicam outros tipos de exigências no que diz respeito aos aspectos qualitativos das inferências, como identificar os defeitos da generalização, em particular aqueles que consistem em, a partir de poucas informações locais, tirar conclusões para o conjunto da população ou universo. Uma segunda exigência consiste em identificar as formas ideológicas que interferem na generalização (p. 37-38)*".

No que tange ao campo de observação, realização de amostragens e representatividade qualitativa, a Pesquisa-Ação proposta por THIOOLLENT (1986), existem três posições relativas à realização de amostras nas pesquisas:

a) Exclusão da pesquisa por amostra, uma vez que a ação coletiva implica em considerar a totalidade dos indivíduos de um grupo;
b) Recomendação do uso da amostra, a partir de critérios estatísticos
c) Valorização de critérios de representatividade qualitativa, uma vez que na prática da pesquisa social a representatividade dos grupos investigados se dá por critérios quantitativos (amostragem estatísticamente controlada) e por critérios qualitativos (interpretativa ou argumentativamente controlados). Isso implica em valorizar as "amostras intencionais" – pequeno grupo de pessoas que são selecionadas de modo intencional em função da relevância que possuem no que tange a um determinado tema ou assunto, o que é sistematicamente aplicado em Pesquisa-Ação, não obstante possa infringir o principio da aleatoriedade que, em geral, é considerado como condição de objetividade (THIOLLENT, 1986, p. 62).

A Pesquisa-Ação implica na constituição de um ou mais grupos de trabalho para realização de suas atividades, envolvendo pesquisadores e participantes. Na abordagem dessa Tese, utilizou-se a técnica do Grupo Focal descrita no tópico 3.4.

3.3 A Pesquisa Experimental e O Projeto de Experimentos

Além da Pesquisa-Ação, utilizou-se no trabalho de pesquisa vinculado a essa tese elementos de pesquisa experimental. Uma vez definido o *software* de descoberta de conhecimento em textos a ser utilizado na Eletronuclear, realizou-se uma série de

experimentos para verificar sua adequação à realidade da organização, sua competência na recuperação de informações e seu real impacto como ferramenta de apoio à gestão do conhecimento.

Um experimento pode ser definido como um teste ou uma série de testes, onde são feitas mudanças propositais nos dados de entrada, de um sistema ou de um processo, de modo que seja possível observar e identificar o efeito dessas mudanças, sobre os dados de saída. O Projeto de Experimentos, por sua vez, diz respeito ao planejamento e condução do experimento, além da análise dos dados de saída, para que conclusões válidas e objetivas possam ser obtidas desse experimento (MONTGOMERY, 1997).

Um Projeto de Experimentos pode consistir de uma série de combinações experimentais de ordem aleatória, onde, em cada combinação, testam-se diferentes ajustes dos fatores de controle e registram-se os resultados observados, os quais devem ser analisados (Montgomery, 1991). Os objetivos principais de análise dos resultados de um projeto experimental são (MONTGOMERY, 1997):

> *(i) Identificar a influência dos fatores controláveis sobre as variáveis de resposta; (ii) Identificar ajustes dos fatores controláveis que resultem em respostas com propriedades desejáveis, tais como próximas a valores nominais e com variabilidade mínima; e (iii) Identificar ajustes dos fatores controláveis que minimizem o efeito de fatores não controláveis (tais como temperatura e umidade no ambiente de experimentação) sobre as variáveis de reposta.*

As etapas principais necessárias para o desenvolvimento de um experimento são as seguintes (RIBEIRO *et al*., 2000):

> *(i) identificação do problema; (ii)planejamento e execução do experimento; (iii) modelagem da variável de resposta e da sua variância; (iv) escolha da função e de critérios a serem utilizados na otimização do experimento; e (v) formalização da otimização.*

No contexto dessa Tese, o Projeto de Experimentos não foi utilizado de forma rígida, conforme previsto em diferentes abordagens teóricas sobre Metodologia de Pesquisa, com medições estatísticas ou de outra ordem matemática. Ele foi desenhado de forma a:

- Complementar e enriquecer o processo desenvolvido sob a égide da Pesquisa-Ação;
- Atuar de modo associado a uma abordagem interativa de desenvolvimento de software;
- Submeter diferentes coleções de documentos textuais digitais ao processamento pelo software escolhido;
- Permitir que os resultados apresentados pelo software pudessem ser avaliados pelo grupo focal constituído para a presente pesquisa, que com seu conhecimento informal, conforme descrito em THIOLENT (1986, p. 67-69), poderia validar ou não tais resultados, do ponto de vista qualitativo.
- Permitir que, as avaliações dos resultados servissem de insumo para o redesenho de rotinas de modelagem e programação do software em uso de modo a adequá-lo às exigências dos resultados a serem obtidos no processo de mineração de textos. Dessa forma, os resultados permitiram ajustar algoritmos, selecionar os melhores recursos de inteligência artificial a serem aplicados, bem como atualizar e aprimorar o desenho dos programas de computador componentes do software em termos de adequação, usabilidade e desempenho.

3.4. Grupo Focal

Os grupos focais (utilizados na Eletronuclear, conforme capitulo 11) têm sido utilizados há cerca de cinqüenta anos em pesquisas qualitativas. Tiveram inicio em processos terapêuticos grupais conduzidos por psiquiatras. Tornou-se, desde então, importante ferramenta em várias áreas, como por exemplo, no campo da Administração Mercadológica. Na concepção de VAUGHN et alli. (1996), grupo focal é instrumento qualitativo que pode ser usada isoladamente ou com outros instrumentos qualitativos ou quantitativos para aprofundar o conhecimento das necessidades de usuários e clientes. Sua meta é compreender o que as pessoas têm a dizer sobre alguma coisa e por quê. De acordo com McDANIEL & GATES (2003) nessa ferramenta:

> “*a ênfase está em fazer com que as pessoas falem à exaustão e detalhadamente sobre o assunto escolhido. A intenção é descobrir o que elas sentem em relação a um produto, um conceito, uma idéia ou uma organização, como tudo isso se*

encaixa na vida delas e seu envolvimento emocional com essas coisas (p.123-124)” .

Para CAPLAN (1990), os grupos focais são

> “*pequenos grupos de pessoas reunidos para avaliar conceitos ou identificar problemas*” (p.527).

O propósito principal de um grupo focal é o de determinar as percepções, os sentimentos, atitudes e idéias dos que dele participam sobre um certo tema ou objeto, que pode ser um produto, proposta, pesquisa, etc.. Em pesquisas de cunho fenomenológico, como é o caso da Pesquisa-Ação, essa técnica se presta a entender como os participantes vêem e interpretam a realidade e quais são seus conhecimentos e experiências. Sua utilização parte do pressuposto que há um elemento de sinergia entre os participantes e que faz com que os resultados ultrapassem as contribuições meramente individuais. O uso desses grupos envolve:

- Definição de um objetivo
- Seleção de um participante
- Seleção de um moderador
- Criação de guia de discussão (esboço dos tópicos que devem ser abordados nos trabalhos do grupo e nas diferentes sessões ou reuniões que realiza) Esse guia é geralmente desenvolvido pelo moderador com base nos objetivos da pesquisa e nas informações necessárias para que determinado objetivo seja atingido McDANIEL & GATES (2003)
- Condução das atividades do grupo em quantas sessões ou reuniões forem necessárias
- Elaboração de um relatório de conclusão das atividades do grupo.

Na ELETRONUCLEAR essa técnica se mostrou adequada e pertinente, afinada com os objetivos da pesquisa da presente tese, uma vez que, além de permitir um conhecimento aprofundado de aspectos operacionais da organização, podemos considerar que a empresa constitui público-alvo da tecnologia em desenvolvimento, sob um processo interativo. Em uma perspectiva de Administração Mercadológica, a ELETRONUCLEAR pode ser vista como cliente para adquirir e utilizar um software de mineração de textos.

3.5 O Processo Interativo de Desenvolvimento de Software e a Orientação a Objetos

A presente Tese serviu-se de um *software* de mineração de textos para implementação prática de suas propostas e avaliação de resultados. Tal *software* poderia ter sido uma ferramenta já existente e disponível no mercado ou desenvolvida. Embora o objetivo principal da Tese não seja o desenvolvimento de um *softwar*e, o processo de desenvolvimento do mesmo, em uma abordagem interativa, contribuiu para uma melhor compreensão das necessidades e da pertinência da proposta para a organização em estudo no que tange à dinâmica da Gestão do Conhecimento na ELETRONUCLEAR, facilitando o entendimento mais amplo dos problemas por ela enfrentados e contribuindo de modo efetivo para a viabilização de adoção de uma estratégia de Pesquisa-Ação.

Um processo interativo de desenvolvimento de software é aquele em que há uma forte participação do usuário em todas as fases, que são continuamente enriquecidas pelos múltiplos e sucessivos *feedbacks* – avaliações, sugestões, críticas e propostas de melhoria – realizadas pelo usuário, permitindo revisões nos estudos dos requisitos e análises necessárias à sua construção. Não se espera o software ficar totalmente pronto para submetê-lo a uma apreciação do usuário, mas isso é feito desde o projeto. Protótipos são então construídos e vão se transformando gradativamente na versão definitiva Em tecnologia da informação existem várias estratégias e metodologias possíveis para uso e processos de desenvolvimento de *softwar*e. O desenvolvimento do *software* ***DifBrain***®, usado nessa Tese, teve como parâmetro a análise e projeto orientado a objeto, cuja essência é enfatizar a consideração de um domínio de problema e de uma solução lógica segundo a perspectiva dos objetos (coisas, conceitos, e entidades) de acordo com LARMAN (2000). Um objeto, ou classe é, de acordo BOOCH, RUMBAUGH & JACOBSON (2000):

> *"alguma coisa geralmente estruturada a partir do vocabulário do espaço-problema ou do espaço da solução; uma classe é a descrição de um conjunto de objetos comuns. Todos os objetos tem uma identidade (você pode atribuir-lhes nomes ou diferencia -los dos demais objetos de alguma maneira), um estado (costuma haver dados a eles associados) e um comportamento (você poderá fazer algo com o objeto ou ele poderá fazer algo com outros objetos) (p.11)".*

No contexto da análise e projeto orientados a objetos, o desenvolvimento do software ***DifBrai***n® seguiu processos e fluxos baseados na área de Engenharia de Software, tendo por base a metodologia do processo unificado[3] *(Unified Process – U*P). Um processo de desenvolvimento de software é um método para organizar as atividades relacionadas com a criação, entrega e manutenção de sistemas de software. De acordo com PAULA FILHO (2003), esse método possui as seguintes características centrais:

- Ser dirigido por casos de uso[4] ;
- Ser centrado na arquitetura;
- Ser interativo e incremental.

Também para esse autor o *UP*, sendo interativo e incremental, constitui um ciclo, em espiral, no qual há uma prototipagem evolutiva, em que há uma série de versões provisórias, as quais, a cada ciclo, cobrem cada vez mais requisitos até que se atinja o produto desejado, permitindo que os requisitos sejam definidos progressivamente e apresentando alta flexibilidade. Cada ciclo é composto por fases, que são as seguintes:

- Concepção – justifica-se a execução de um projeto de desenvolvimento de software do ponto de vista de uma atividade a ser executada ou, comercialmente, de um negócio de um cliente;
- Elaboração – o produto é detalhado o suficiente para permitir um planejamento acurado da fase de construção;
- Construção – é produzida uma versão operacional do software;
- Transição – o produto é colocado à disposição de uma comunidade de usuários.

---

[3] O processo unificado, abordagem consolidada para desenvolvimento de sistemas, foi proposto em JACOBSON, I. RUMBAUGH, J. BOOCH, G. *Unified Software Development Proces*s. Addison-Wesley, Reading – MA. , 1999. Não faz parte do escopo do trabalho discutir tal metodologia de desenvolvimento, já consolidada na área de tecnologia da informação. Apresentamos apenas as informações fundamentais. De acordo com PAULA FILHO (2003), um processo é um conjunto de passos parcialmente ordenados para se atingir a uma meta.

[4] Um caso de uso é uma descrição narrativa de uma atividade de um domínio, como, por exemplo, emprestar livros em uma biblioteca. Um ciclo de desenvolvimento é a implementação de um ou mais casos de uso, ou de versões simplificadas de casos de uso. Os casos de uso constituem um parte dos diagramas da Linguagem de Modelada Unificada – UML (Unifiel Modeling Language) uma notação sobretudo diagramática para modelagem de sistemas, usando conceitos orientados a objetos. Além dos casos de uso, a UML define os diagramas de classes, objetos, seqüências, colaborações, gráficos de estados, atividades, componentes e implantação (BOOCH, RUMBAUGH & JACOBSON, 2000).

Além disso, em cada fase as atividades possuem fluxos de trabalho composto por:

- Estudo de Requisitos – fluxo que visa obter um conjunto de requisitos de um produto, acordado entre usuário e desenvolvedor;
- Análise – fluxo que objetiva detalhar, estruturar e validar os requisitos de forma a permitir planejamento detalhado.
- Desenho – fluxo que permite formular um modelo estrutural do produto que sirva de base para implementação.
- Implementação – fluxo que visa realizar o desenho em termos de componentes de código.
- Testes – Fluxo cujo objetivo é verificar o resultado da implementação.

3.6. Considerações Finais Sobre o Emprego de Metodologia

O presente trabalho é resultado de um processo longo de interação entre o pesquisador autor da presente Tese e a organização pesquisada: a ELETRONUCLEAR. Esse processo de interação deu-se em muitas visitas às instalações da empresa, em Angra dos Reis, troca de correspondências eletrônicas com seus funcionários, acesso a documentos, conversas formais e informais com seus funcionários e interlocução com um responsável principal pelo projeto prospectivo de uso de mineração de textos na ELETRONUCLEAR. Através de sucessivas análises, apresentação de resultados, análise do *feedback* dos responsáveis, foi possível implementar o trabalho.

O desenvolvimento da pesquisa que levou aos resultados do presente trabalho após um período longo de atuação em que houve forte interação do autor com uma equipe de profissionais da ELETRONUCLEAR e com os quais passou-se a buscar solução para problemas complexos relativos à geração e entrega de informações estratégicas a usuários especiais que executam tarefas relativas à operação e manutenção de usinas nucleoelétricas. Por isso, a abordagem da Pesquisa-Ação mostrou-se indicada ao processo, associando-se a ela um projeto de experimentos com *software* de mineração de textos específico

PARTE II – REFERENCIAL TEÓRICO

4 GESTÃO DO CONHECIMENTO

4.1 Aspectos Preliminares e Definições

Não é fácil definirmos conhecimento, como discutiremos a seguir. O fato concreto é que conhecimento humano, que sempre foi importante, tornou-se, a partir da segunda metade do século XX, elemento decisivo na condução de mudanças econômicas, sociais e políticas da humanidade, passando a ser cada vez mais elemento de transformação e diferenciação crescente entre pessoas, grupos sociais e países. O desenvolvimento científico e tecnológico intensificou, a partir dessa época, a imposição de mudanças e transformações à sociedade e às organizações em um ritmo cada vez mais acelerado. O conhecimento começou a ser visto como um emergente fator de produção, somando-se aos tradicionais fatores terra, trabalho e capital. O acesso e o domínio do conhecimento foi se caracterizando cada vez mais como vantagem competitiva, seja de empresas, seja de nações. A capacidade de adicionar valor agregado aos produtos, a partir do resultado do desenvolvimento científico, tecnológico ou da melhoria de processos organizacionais passou a fazer cada vez mais a diferença no sucesso de organizações e países. Em seu famoso artigo *"Como nós pensamos"*, BUSH (1945) afirmou:

> *"O conhecimento humano vem crescendo assustadoramente e se tornando cada vez mais especializado. À medida que o conhecimento cresce e a especialização se estende, mais complicado se torna o nosso acesso a esse monumental acervo. O investigador fica perplexo quando tem que lidar com o produto da pesquisa de milhares de colegas -- não tendo tempo para ler, muito menos para analisar e memorizar, tudo o que é publicado, mesmo em sua área de especialização. A especialização talvez seja a única maneira de fazer o conhecimento progredir. Mas o esforço de construir pontes entre as várias disciplinas especializadas ainda é incipiente e artificial"*

A gestão do conhecimento nasceu do esforço organizacional em tratar de forma intencional e sistematizada a questão do conhecimento de forma a desenvolver um conjunto de possibilidades de sobrevivência em cenários de turbulência e alta competição, bem como de gerar crescimento a partir da construção de vantagens competitivas apoiadas no conhecimento. De acordo com NEVES e NASCIMENTO (1999):

> *"as pessoas sempre detiveram conhecimento, adquirido através de informações e experiências. O que as organizações estão descobrindo agora são maneiras*

de transformar e gerir esse conhecimento existente para alavancar vantagens competitivas e gerar novos conhecimentos, auxiliadas ou não por novas tecnologias".

4.1.1 Dados, Informação, Documento e Conhecimento

Em muitos escritos de Ciência da Informação há uma proliferação de usos dos termos dados, informação e conhecimento, sendo que muitas vezes esses termos ou são tratados como sinônimos ou no fluxo das discussões e análises há uma passagem quase automática de uma instância à outra. O problema é que o caminho dados => informação => conhecimento => gestão do conhecimento não se faz de modo tão automático, havendo diferenças entre tais instâncias e um processo humano e social quanto mais se caminha em direção ao conhecimento. A definição e diferenciação entre o que venha a ser dados informação ou conhecimento é polêmica e não há a intenção do presente trabalho em exaurir tal questão. Entretanto, para o fluxo da discussão apresentada nessa tese, torna-se necessário realizar uma caracterização de referência para o contexto da discussão sobre o assunto que é aqui tratado. DAVENPORT & PRUSAK são autores que se preocuparam com a diferenciação de tais conceitos. Na seqüência, será feito um exame mais detalhado sobre a questão da gestão do conhecimento. A Figura 2 a seguir pretende expressar inicialmente a evolução dado => informação => conhecimento => sabedoria (entendida aqui como conhecimento em que se adicionam valores e uma postura ética). As setas verticais indicam os elementos que vão sendo adicionados a cada instância, que apreciaremos a seguir.

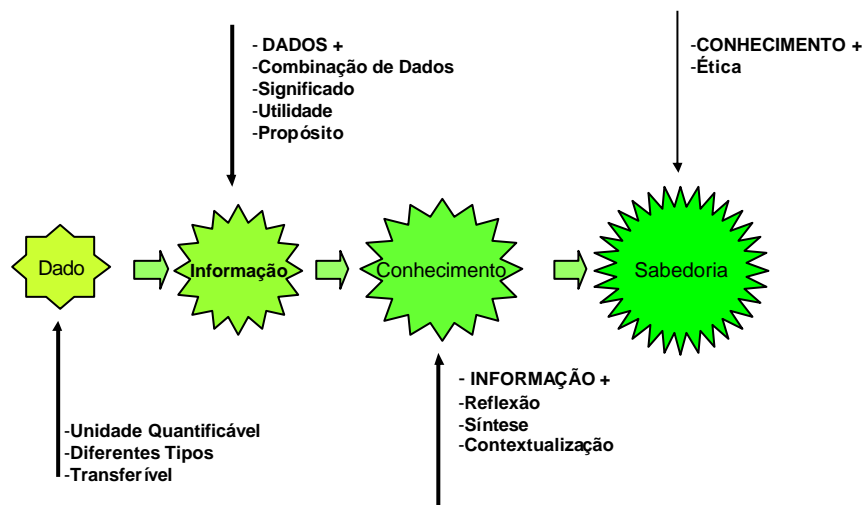



Figura 2: Dos dados à sabedoria. Fonte: Elaborado pelo autor.

4.1.1.1 Dados

De acordo com SETZER (1999):

> “*Um dado é puramente objetivo – não depende do seu usuário. A informação é objetiva-subjetiva no sentido que é descrita de uma forma objetiva (textos, figuras, etc), mas seu significado é subjetivo, dependendo do usuário. O conhecimento é puramente subjetivo – cada um tem a experiência de algo de forma diferente*”(p.3).

Para LE COADIC (2004), dados são representações convencionadas, codificadas de uma informação em uma forma que permita submetê-la a processamento eletrônico em que as letras das palavras são representadas por dígitos binários (0 ou1) no código ASCII que expressa um conjunto de sinais elétricos. Na perspectiva de TURBAN (2003), dados são fatos puros ou descrições básicas de coisas, eventos, atividades e transações que são capturados, registrados, armazenados e classificados, mas não são organizados para transmitir qualquer significado por si só. Dado é a matéria-prima da informação. Na visão de REZENDE (2003) é um elemento puro, quantificável sobre um determinado evento e, em geral, usados em ambiente operacional, registrados, selecionados e recuperados de um banco de dados ou de diferentes formas de documentos. De acordo com SOUZA (2006):

> “*dados podem ser definidos como seqüências de símbolos para os quais são atribuídos significados; símbolos estes que podem ser codificados, interpretados e manipulados por programas de computador, e enviados através de redes e dispositivos de comunicação*”.

Em geral, um dado por si só não confere sustentação ao entendimento de uma situação. Sob um ponto de vista mais simples, de uma lógica computacional, um “dado” é tudo aquilo que é fornecido como entrada para um processamento e informação, é tudo aquilo que o processamento gera como uma saída. De acordo com DAVENPORT & PRUSAK (1998):

> “*dados são um conjunto de fatos distintos e objetivos relativos a eventos. Num contexto organizacional, dados são utilitariamente descritos como registros estruturados de transações*”.

4.1.1.2 Informação

Sob o ponto de vista etimológico, conforme ZEMAN (1970), o termo "informação" é originário do latim, derivando do verbo *informare*, que significa dar forma, colocar em forma mas também representar uma idéia ou noção. Informações podem ser entendidas como dados contextualizados, objetivos, confiáveis e utilizáveis para algum propósito. As informações nascem de dados processados. A informação envolve a interpretação de um conjunto de dados, o qual só faz sentido em um contexto de referência, ou seja, da informação (REZENDE, 2003). Podemos defini-la como um conjunto de fatos (dados) organizado de modo a fazer sentido para o destinatário (usuário). Para O'BRIEN (2001), a informação requer a elaboração de cenário, a identificação dos elementos que atuam nesse cenário, a localização do dado nesse contexto e a avaliação das relações entre o dado em questão e ambiente onde ele está inserido. Para esse autor, a informação é o dado contextualizado. Portanto, a manipulação do dado através de modelos que o contextualizem constitui uma forma de construir informações, e estas são fundamentais na gestão do conhecimento ou em qualquer atividade humana. De acordo com TURBAN (2003), para ser útil ao usuário, a informação precisa ser completa, flexível, confiável, relevante, pontual, verificável, acessível e segura.

Na verdade, temos problemas para definir o que seja informação. Conforme LANCASTER (1989):

> *Informação é uma palavra usada com freqüência no linguajar quotidiano e a maior parte das pessoas que a usam pensam que sabem o que ela significa. No entanto, é extremamente difícil definir informação, e até mesmo obter consenso sobre como deveria ser definida. O fato é, naturalmente, que informação significa coisas diferentes para pessoas diferentes.*

Segundo KANDO apud MACHADO (2001) é possível identificar três grupos distintos acerca do conceito de informação para a Ciência da Informação:

> *a) "Informação com entidade objetiva: compreende o conteúdo do documento;*
> *b) "Informação com entidade subjetiva: representada pela imagem-estrutura do receptor e as permutas da mesma;*
> *c) "Informação com processo: faz referência ao processo mediante o qual o sujeito se informa".*

Na perspectiva de SHANNON (1998) informação é algo que um receptor recebe de um emissor em processo de comunicação, ou seja, em uma mensagem. Entretanto, tal mensagem requer ser devidamente compreendida para, posteriormente ser reconhecida como contendo algo de novo para o receptor. Quando um receptor recebe uma mensagem contendo alguma coisa que já ele já entende ou então que não compreende, essa mensagem não transmite informação para ele. Assim sendo, a idéia de que dados ou mensagens podem trazer informação é uma coisa subjetiva. Alguns dados podem gerar informações para um determinado usuário e não gerar para outros, além de poderem trazer diferentes informações, em diferentes quantidades e em diferentes tempos. É possível também que dados diferentes possam vir a transmitir uma mesma informação e na verdade tudo dependerá daquele que recebe e analisa os dados com o estado de conhecimento que possui. Para SOUZA (2006) o conceito de informação é mais abstrato, sendo que a informação não prescinde do sujeito que a depreenda a partir dos dados, no ato conhecido como interpretação. De acordo com DAVENPORT & PRUSAK (1998):

> *a informação é descrita como uma mensagem que tem sempre um emissor e um receptor e que tem por finalidade mudar o modo como o destinatário vê algo, exercer algum impacto sobre seu julgamento e comportamento. (...) Os computadores podem ajudar a agregar os dados e a transformá-los em informação, porém quase nunca eles ajudam na parte de contexto, e os seres humanos geralmente precisam agir nas partes de categorização, cálculo e condensação. (...)*

MESSIAS E MORAES (2003) avaliam que o conceito de informação não pode ser desenvolvido isoladamente, o que implica na união de diversas abordagens e perspectivas teórica, sendo possível classificar algumas dessas abordagens de forma sintetizada, tais como centrada na mensagem, na tomada de decisões, nas estruturas semióticas, no significado, no processo e na cognição.[5]

---

[5] "A abordagem da informação centrada na mensagem ou teoria matemática da informação privilegia a teoria de Shannon & Weaver, que descreve o funcionamento de um sistema mecânico, onde as mensagens emitidas pela fonte são transmitidas por um canal a fim de serem recebidas com o mínimo de deformação por um usuário. Nesse contexto, a importância está centrada no canal e na sua capacidade de veicular mensagens a um custo baixo. A abordagem pragmática pressupõe que a informação é um elemento que auxilia na tomada de decisão de um sujeito. Assim sendo, a mensagem funciona com um redutor de incertezas. Na abordagem estruturalista, a informação é vista como estruturas semióticas, caracterizada por uma estruturação deliberada da mensagem pelo emissor, com o objetivo de atingir a estrutura da imagem do receptor. Sendo a informação modificadora de estruturas cognitivas. A abordagem centrada no significado é oriunda da lingüística e centra-se na organização da mensagem em três níveis: predominância sintática, semântica ou pragmática. A abordagem centrada no processo considera a informação como um processo que ocorre na mente humana quando um problema e dado útil para sua

Para ARAÚJO (1994) a informação é algo dinâmico, que faz com que o agente modifique seu estado de conhecimento do mundo atual, modificando estruturas, e só tendo sentido no contexto da ação acabada, ou seja, no contato efetivo de uma mensagem, um potencial de infomação e o usuário.

4.1.1.3 Conhecimento

Na perspectiva de MIZZARO (1996), conhecimento é a forma que a pessoa percebe o mundo, processando dados e informações que recebe e transformando-os em conhecimento, o qual é um potencial para um conjunto de ações. Esse conhecimento é mutável e, ainda conforme esse autor, o conhecimento de uma pessoa em determinado momento é o seu estado de conhecimento. Conhecimento envolve experiência, julgamentos e correlações entre informações e contextos de interesse do usuário. De acordo com REZENDE (2003), conhecimento diz respeito à habilidade de criar um modelo mental que descreva o objeto e indique as ações a implementar e as decisões a tomar. A compreensão, a análise e síntese, necessárias para a tomada de decisões inteligentes, são realizadas a partir do nível do conhecimento. Em DAVENPORT & PRUSAK (1998), as relações entre dados, informações e conhecimento são relevantes e, em geral, essas categorias podem ser relativamente bem separadas. Esses autores afirmam que:

> "*conhecimento não é dado nem informação, embora esteja intimamente relacionado com ambos e que o sucesso ou fracasso organizacional muitas vezes pode depender de se saber de qual deles precisamos, com qual deles contamos e que podemos ou não fazer com cada um deles* (DAVENPORT & PRUSAK, 1998: 1).".

---

solução encontram-se numa união produtiva. E, finalmente, a abordagem cognitivista apresenta a informação e o conhecimento como elementos diferentes, sendo o conhecimento avaliável somente em nível mental e a informação sendo o substituto físico usado para a comunicação. Por meio dessa reflexão podemos perceber que o conceito de informação pode ser desenvolvido em ambientes diversos, englobando vários ramos do saber, tais como, comunicação, filosofia, lingüística, teoria matemática da informação e até na biologia, que compreende a informação enquanto regularidades mantidas em sistemas complexos. (MESSIAS, 2002). É importante ressaltar que essa variedade de abordagens ao termo informação dificulta em sua compreensão. Na ausência de delimitações precisas, muitos outros termos são utilizados como sinônimo de informação, a começar por dado e conhecimento. Essa discussão é muito antiga e muito extensa. Nosso objetivo nesse momento não é aprofundar essa questão." MESSIAS E MORAES (2003, p. 6-7).

Esses mesmos autores tentam estabelecer uma diferença entre os três conceitos com objetivo de melhor situar o tema do conhecimento empresarial. Na sua abordagem:

> *O conhecimento é uma mistura fluida de experiência condensada, valores, informação contextual e insight experimentado, a qual proporciona uma estrutura para a avaliação e incorporação de novas experiências e informações. Ele tem origem e é aplicado na mente dos conhecedores. Nas organizações, ele costuma estar embutido não só em documentos ou repositórios, mas também em rotinas, processos, práticas e normas organizacionais* (DAVENPORT & PRUSAK, 1998: 2-6)."

Sob o ponto de vista organizacional, conhecimento consiste em informações organizadas e processadas para transmitir discernimento, experiência, aprendizagem acumulada ou habilidade, se aplicável a um problema ou processo empresarial atual. Enfim, de acordo com BURKE (2003), conhecimento é o que foi processado pelo pensamento e existem diferentes tipos de conhecimento em toda cultura. Há o conhecimento filosófico, o senso comum, o científico, o religioso, dentre outros e todos geram objetos pelos quais são representados.

4.1.1.4 Documento

Na presente tese, que aborda a mineração de textos em documentos, faz-se importante abordar o conceito de documento. O documento pode ser visto, baseando-se em BRIET (1951) [6] uma evidência que suporta um fato, um signo de natureza física ou simbólica que permite representar, preservar, registrar e reconstituir ou demonstrar um fenômeno físico ou conceitual.

De acordo com MESSIAS E MORAES (2003):

> *"A Ciência da Informação parece privilegiar a visão de informação como conhecimento (de alguma forma) registrado, atrelado ao conceito de documento, e a transferência de tais conteúdos informacionais seria o foco de maior interesse da área. Podemos constatar também que o conceito de informação para a Ciência da Informação mantém uma relação muito estreita com o conceito de documento e consequentemente com os processos realizados na instituição biblioteca" (p. 11).*

---

[6] Citado em MESSIAS E MORAES (2003)

Em geral, para a Biblioteconomia e a Ciência da Informação, o documento é a unidade básica que viabiliza expressão de um determinado conhecimento, através das informações nele contidas, sendo, de maior interesse, a transferência de conteúdos através de tais documentos ou objetos portadores de informação. Por outro lado, BRAGA e CHRISTOVÃO (1997) postulam que documentos não são e nem contém informação. Documentos são mensagens que podem ou não produzir informação, dependendo do estado cognóscio do receptor.

Na perspectiva de Paul Otlet, no final do século XIX, conforme citado por OLIVEIRA (1) (2005) documento passou a ser "o livro, a revista, o jornal, a peça de arquivo, a estampa, a fotografia, a medalha, a música, o disco, o filme e toda parte documentária que precede ou sucede a emissão radiofônica". Segundo LE COADIC (2004),

> *"documento é um termo genérico que designa os objetos portadores de informação. Um documento é todo artefato que representa ou expressa um objeto, uma idéia, ou uma informação por meio de símbolos gráficos e icônicos (palavras, imagens, diagramas, mapas, figuras, símbolos), sonoros e visuais (gravados em suporte de papel ou eletrônico). O documento, segundo o tipo de suporte, é denominado documento de papel ou documento eletrônico" (LE COADIC, 2004:5* .

Em uma perpectiva mais ampla, de acordo com BUCKLAND (1991), um documento inclui também imagens, sons ou qualquer outro objeto informativo e não apenas os textos, que obviamente tem um papel de destaque, especialmente nas bibliotecas digitais atuais. Em outro trabalho BUCKLAND (1997) afirma que documento é qualquer forma ou expressão que o pensamento humano possa vir a assumir ou qualquer fonte de informação, materialmente expressa que possa ser usada como referência ou estudo como uma autoridade. De acordo com ele, o que confere a condição de documento a um determinado objeto ou a um ser é a possibilidade de ser relacionado com outras evidências.

No presente trabalho, é importante delimitar o conceito de documento, uma vez que será discutido o conceito de mineração de textos. Nesse contexto, portanto,

mantém-se para foco de análise a idéia de documento como um objeto de informação textual, portador de assuntos, conceitos, autoria, local e data de elaboração, e passível de interesse de uma determinada comunidade de usuários potenciais. Sendo descritos em forma de texto, portanto, estão em escritos em linguagem natural e como tal devem ser alvo de tratamento pela mineração de textos.

4.1.2 A Questão da Gestão do Conhecimento

A questão do conhecimento cresceu em relevância para o campo da Administração no seu esforço para gerenciar e desenvolver as organizações, muito ligada à gestão do fator humano e de como responder aos desafios de como as pessoas podem interagir e compartilhar experiências e informações importantes para o bom manejo e operação dessas mesmas organizações. As questões da obtenção, geração, tratamento e uso da informação que suportam os múltiplos relacionamentos das pessoas na organização são uma derivada desse processo. Preocupações sobre o que seja conhecimento e os seus diferentes tipos, bem como quais são interessantes para as organizações também aparecem. Há também críticas ao conceito. Por isso, o campo da Gestão do Conhecimento é interdisciplinar, envolvendo a Administração, a Ciência da Informação, a Ciência da Computação, a Pedagogia, a Psicologia e a Filosofia.

O conceito de gestão do conhecimento não possui um consenso e pode ser considerado e estudado sob diversos prismas. Inicialmente, precisamos colocar foco numa reflexão sobre o que venha a ser conhecimento e, em seguida, sobre os desdobramentos sobre as questões organizacionais. Nesse sentido, podemos tratar o assunto a partir da consideração de quatro prismas: a) filosófico-cognitivo; b) organizacional; c) informação e tecnologia e d) fator humano. É possível, também, fazer uma reflexão crítica sobre a evolução da gestão do conhecimento. Ao final, o enfoque principal recairá sobre questões organizacionais e tecnológicas, discutindo contribuições de autores que foram considerados importantes para a consolidação de uma visão mais articulada sobre gestão do conhecimento organizacional, as quais se tornaram importantes para efeito desse trabalho, em especial as de POLANYI (1966) que foram absorvidas por NONAKA & TAKEUCHI (1997) e DAVENPORT & PRUSAK (1998) e CHOO (2003).

4.2. Prisma Filosófico-Cognitivo

4.2.1 Face Filosófica

Podemos considerar que a filosofia e a epistemologia sempre estiveram envolvidas na questão do entendimento do que seja conhecimento e de suas possíveis conceituações, as quais podem ser diversas e não pretendem esgotar-se nesse trabalho. O próprio conceito de epistemologia pode apresentar variações. Classicamente, conforme visão de GARCIA MARCO (1993) a epistemologia pode ser entendida como a parte da filosofia voltada ao estudo dos processos do conhecimento humano, considerando sua lógica, seus princípios e suas origens. Porém, hoje[7] a esfera de estudo da epistemologia se concentra em uma perspectiva não tão ampla, focando o estudo das características do discurso científico e da evolução dos paradigmas da ciência dentro de um campo de conhecimento.

Conhecimento, derivado do latim, *cognoscere*, que quer dizer conhecer ou procurar, indicando a função ou ato da vida psíquica que tem por efeito tornar os objetos presentes ao sentido ou à inteligência. É algo, portanto, colocado no nível do humano, do pessoal, embora possa receber a moldura da cultura, da história, da ética e dos padrões de desenvolvimento tecnológico de uma determinada sociedade em um determinado momento do espaço e do tempo. Podemos citar alguns marcos importantes que envolvem a abordagem da filosofia e da epistemologia ao conhecimento. Na Grécia antiga a questão do conhecimento teve um papel central. A expressão filosofia, com o significado de amor à sabedoria, ou de aspiração ao conhecimento já confere ao conhecimento humano um destaque relevante em sua intenção e essência. Em busca da "essência" da filosofia, DILTHEY (1921) afirma que "deve ser primeiramente buscado um conteúdo comum nos sistemas em que se forma a representação geral da filosofia". [8] HESSEN (1999) irá afirmar que

> *"esses sistemas existem, como o são os sistemas de Platão, Aristóteles, Descartes, Leibniz, Kant e Hegel e que quando nos aprofundamos neles, deparamos nos com características essenciais comuns, apesar de todas as*

---

[7] A Epistemologia ficará mais voltada ao estudo das ciências. Ao longo do século XX, três grandes perspectivas podem ser identificadas na Epistemologia: o Empirismo Lógico (Círculo de Viena), o Falseabilismo (Karl Popper) e a Estrutura das Revoluções Científicas (Kuhn).

[8] Cf. DILTHEY, W. Das Wesen der Philosophie, in: Kultur der Gerenwart I, VI: 3ª. ed. 1921.

> *diferenças que apresentam. Encontramos uma atração pelo todo um direcionamento para a totalidade dos objetos. Esses sistemas, portanto, possuem o caráter da universalidade. A essa característica essencial comum soma-se outra. A atitude do filosofo com relação à totalidade dos objetos é uma atitude intelectual, uma atitude de pensamento. Cabe ao filósofo conhecer, saber. O filósofo é um conhecedor por natureza. Aparecem , portanto, as seguintes marcas da essência da filosofia: 1) a atitude em relação à totalidade dos objetos; 2) o caráter racional, cognoscitivo dessa atitude* (HESSEN, 1999:5)".

Ainda segundo HESSEN (1999), dois outros elementos podem ser considerados: a visão de si e a visão do mundo. Há uma interação entre esses quatro elementos. Conforme afirma HESSEN (1999):

> "*a totalidade dos objetos pode se referir tanto ao mundo exterior quanto ao mundo interior, tanto ao macrocosmo quanto microcosmo. Se a consciência filosófica dirige-se ao macrocosmo, tratamos de filosofia como visão de mundo. Se é o microcosmo que constitui o objeto do enfoque filosófico, surge um segundo tipo de filosofia: a filosofia como visão de si, do espírito* (HESSEN, 1999:9)".

HESSEN afirma que é então possível definir a essência da filosofia como

> "*auto-reflexão do espírito sobre seu comportamento valorativo teórico e prático e, igualmente, aspiração a uma inteligência das conexões últimas das coisas ,a uma visão racional de mundo. Podemos, porém, estabelecer uma ligação mais profunda entre esses dois elementos essenciais. Como Platão e Kant nos mostram, existe entre ambos uma relação de meio e fim. A auto-reflexão do espírito é meio para se atingir uma imagem de mundo, uma visão metafísica de mundo. Em conclusão portanto, podemos dizer que a filosofia é a tentativa do espírito humano de atingir uma visão de mundo, mediante a auto-reflexão sobre suas funções valorativas, teóricas e práticas* (HESSEN, 1999:9)".

Um quadro sinóptico poderia apresentar essas idéias de HESSEN (1999) na Figura 3:

	MUNDO EXTERIOR	MUNDO INTERIOR
TOTALIDADE DOS OBJETOS	MACROCOSMO	MICROCOSMO
COGNOSCITIVIDADE	VISÃO DO MUNDO	VISÃO DE SI

Figura 3: Sinopse das Idéias de Hessen sobre a Essência da Filosofia
Fonte: Baseado em HESSEN (1999)

Platão e Aristóteles apresentarão visões diferentes em seus sistemas filosóficos. Para HESSEN (1999) a visão de Platão significa que a filosofia busca uma *visão de si, do espírito,* estendendo-se à totalidade dos objetos, sejam eles práticos, sejam eles coisas como virtude, valores do bom, do belo e do verdadeiro. Já em Aristóteles temos uma concentração no que podemos chamar de *conhecimento científico e em seu objeto, o ser.*

> *"No seu núcleo está uma ciência universal do ser:"a filosofia primeira" ou como seria chamada mais tarde, a metafísica. Ela nos informa sobre a essência das coisas, a contingência e os princípios últimos da realidade. Se a filosofia socrático-platônica pode ser caracterizada como uma visão de si do espírito, devemos dizer que, em Aristóteles, a filosofia aparece antes de tudo como visão de mundo* (HESSEN, 1999:6)"

NONAKA E TAKEUCHI (1997) Na discussão sobre o que seja conhecimento, irão afirmar:

> *"Apesar das diferenças fundamentais entre racionalismo e empirismo, os filósofos ocidentais, em geral, concordam que conhecimento é a "crença verdadeira justificada", um conceito introduzido inicialmente por Platão em Ménon, Pédon, e Teeteto . No entanto, a definição de conhecimento está longe de ser perfeita em termos lógicos. Segundo essa definição, nossa crença na verdade de uma coisa não constitui nosso verdadeiro conhecimento dessa coisa, por isso existe uma chance, por menor que seja, de que nossa crença esteja errada. Portanto, a busca do conhecimento na filosofia ocidental é carregada de ceticismo, o que induziu diversos filósofos a buscarem um método que os ajudasse a estabelecer a verdade indubitável do conhecimento. Eles almejavam descobrir o "conhecimento fundamental sem prova ou indício" sobre o qual seria possível assentar todo e qualquer conhecimento.* (NONAKA E TAKEUCHI,1997: 24-25)".

Esses autores lembram, portanto, que o ocidente é marcado por duas grandes tradições epistemológicas: o racionalismo e o empirismo. Afirmam que no racionalismo, marcado pelos sistemas de Sócrates, Platão, Descartes e Kant, o conhecimento é produto de um processo mental ideal, enquanto que para o empirismo, marcado pelos sistemas de Aristóteles, Locke, Espinosa e Leibniz, o conhecimento é produto da experiência sensorial. Lembram que também no método há distinções importantes: enquanto o racionalismo utiliza a *dedução*, que recorre a construções mentais, como conceitos, leis ou teorias, o empirismo recorrerá à *indução,* servindo-se de experiências sensoriais específicas. De qualquer forma, um ponto importante aqui precisa ser destacado: ao lidar com o conhecimento será necessário lidar com a idéia der representação desse conhecimento. Na medida em que os seres humanos conhecem a

realidade, tem a necessidade de assimilar as coisas através das representações na mente da pessoa que conhece.

Outro ponto importante nessa consideração preliminar sobre a questão do conhecimento é a diferença que NONAKA E TAKEUCHI fazem sobre a visão ocidental e oriental (japonesa) do conhecimento. Enquanto na filosofia ocidental, há uma longa tradição de separação entre o sujeito que conhece e o objeto conhecido, mente e corpo ou mente e matéria, na filosofia oriental japonesa postula-se uma unidade do homem com a natureza, do corpo com a mente e, por conseguinte, do eu e do outro. Segundo os autores:

> "*o dualismo cartesiano entre sujeito e objeto ou mente e corpo partia do pressuposto de que a essência de um ser humano é o seu pensante racional. Esse eu pensante busca o conhecimento isolando-se do resto do mundo e dos outros seres humanos. Mas os desafios contemporâneos impostos ao dualismo cartesiano enfatizam a importância de alguma forma de interação entre o eu e o mundo externo na busca do conhecimento* (NONAKA E TAKEUCHI, 1997: 29)".

Na verdade, as afirmações de NONAKA e TAKEUCHI relativas a Kant precisam ser revistas, uma vez que Kant avalia criticamente os contrapontos tanto do racionalismo como do empirismo. Ao propor sua filosofia, Kant o faz considerando duas ameaças: o dogmatismo e o empirismo. O primeiro confia demasiado na razão sem desconfiar o suficiente das ilusões especulativas. O segundo, por medo de erros dogmáticos, tende a reduzir tudo à experiência. Para Kant, a razão precisa ser alvo de um justo uso, após fazer uma triagem daquilo que lhe é possível e daquilo que lhe escapa. Na *Crítica da Razão Pura*[9], investiga os limites do emprego da razão no conhecimento e procura estabelecer as condições de possibilidade do conhecimento, distinguindo os usos legítimos da razão na produção de conhecimento dos usos especulativos que não agregam conhecimento. Para Kant o conhecimento vem da sensibilidade e do entendimento. Pela primeira, os objetos são dados ao homem e pelo segundo são pensados. Sem a conjugação de ambos não é possível a experiência do real. Portanto, é preciso analisar as condições de possibilidade do conhecimento, isto é, refletir criticamente sobre os fundamentos da ciência e da experiência em geral.

---

[9] *Kritik der reinen Vernunft* – editado inicialmente em 1781 e posteriormente em 1787 com alterações.

Assim sendo, sob a face filosófica, a questão do conhecimento pode ser amplamente abordada, sob diferentes perspectivas e múltiplos pontos de vista filosóficos de variados autores, não sendo este o propósito do presente trabalho. Mas, conforme debatido acima, precisamos buscar pontos em comum em tais discussões e fica aqui assinalada a idéia de se buscar pontos em comum em diferentes sistemas filosóficos, sendo estes a busca da universalidade, a atração pela totalidade na compreensão dos objetos do mundo.

4.2.2 Face Cognitiva

4.2.2.1 Behaviorismo e *Gestalt*

O embate entre empirismo e racionalismo se tornou presente na reflexão contemporânea sobre aprendizagem e conhecimento, sendo o primeiro representado pela teoria *behaviorista* (de conotação positivista e mecanicista) e o segundo pelo movimento chamado *gestalt* (de conotação idealista)[10]. O *behaviorismo* assumiu o valor da experiência como determinante do conhecimento o qual é fruto da aprendizagem. Essa aprendizagem implica em condicionamentos, isto é: o indivíduo aprende aquilo para o qual é estimulado e reforçado. Portanto, o comportamento, à luz do behaviorismo, é modelado e isso implica na importância do estudo dos elementos de reforço. A aprendizagem seria, portanto, uma mudança de comportamento que é resultante do treino e da experiência, podendo ser acelerada e direcionada por mecanismos de reforço (estímulos). Para o behaviorismo o conhecimento tem uma conotação associativa, isto é, se constitui numa cadeia de idéias atomisticamente formadas a partir do registro dos fatos que vão sendo acumulados pela experiência do sujeito. Na perspectiva associativa que confere base a essa corrente de pensamento aprender é um processo gradativo pautado em ensaios e erros e no exercício repetitivo.

---

[10] Não é objetivo do presente trabalho analisar em detalhes o *behaviorismo* e *gestaltismo*. A presente explanação sobre *behaviorismo* e *gestaltismo* foi baseada na interpretação de A. GIUSTA Concepções de Aprendizagem e Práticas Pedagógicas artigo publicado na Revista de Educação, Belo Horizonte (1985: 24-31). É um panorama reduzido de seus aspectos e características para contextualizar o estudo do conhecimento e abrir perspectiva para as idéias de J.PIAGET que serão apresentadas a seguir. Uma visão mais detalhada das duas correntes implica na avaliação dos estudos clássicos de J.B. WATSON, I.P.PAVLOV e B.F.SKINNER, dentre outros representantes do *behaviorismo* e de M.WERTHEIMER, K.KOFFKA e W.KÖHLER representantes do *gestaltismo*.

O *gestaltismo*, (do alemão *gestalten*, isto é configuração) por sua vez, caminha no sentido oposto ao do behaviorismo, numa perspectiva racional e subjetiva, afirmando que o conhecimento é anterior à experiência. É resultado do exercício de estruturas racionais pré-formadas no sujeito. Há estruturas mentais como totalidades organizadas e pré-formadas que não são, portanto, fruto da ação desse sujeito. A organização perceptual é o elemento primordial. Perceber totalidades, formas através de "intravisões" (*Einsicht* em alemão que redundou no termo inglês *insight*), ou seja, da reestruturação do campo perceptual. Isso permite que as totalidades sejam melhor captadas da realidade e, uma vez que isso aconteceu, não há mais necessidade de repetições.

4.2.2.2 O Desenvolvimento do Conhecimento na Perspectiva Construtivista

Uma perspectiva diferente das duas anteriores é aberta por PIAGET[11]. Enquanto epistemólogo, PIAGET (1976) ocupou-se significativamente do estudo da formação e desenvolvimento do conhecimento. Analisando sua obra, CELLÉRIER (1978) postula que a problemática inicial daquele autor é fundamentalmente filosófica e nelas se colocam questões clássicas da filosofia, tais como: o que é conhecimento? Como conhecemos? Como conseguimos conhecer o que conhecemos? A filosofia possui uma formulação clássica para tratar essas questões em termos de verdades normativas (que são independentes de verificação empírica, como a lógica e a matemática) e verdades de fato (que são verificáveis empiricamente, como, por exemplo, a solidificação da água a 0^{o} C.). Teríamos então dois tipos de conhecimento, o normativo e o verificativo. Além disso, outra questão se coloca: como estabelecer relações entre ambos? Dar uma resposta satisfatória a este problema da relação existente é o desafio a qualquer teoria do conhecimento e PIAGET reformulará todas essas questões recolocando as perguntas em outros termos, tais como: como se passa de um estado de menor conhecimento a outro

---

[11] Também no que tange à extensa e complexa obra de PIAGET o presente trabalho pretende apenas apresentar uma visão panorâmica geral resgatando conceitos considerados fundamentais para os objetivos do presente estudo. Uma análise mais detalhada deve recorrer, além das inúmeras obras no campo da ciência cognitiva, aos trabalhos e publicações originais do autor, em especial: PIAGET, Jean. <u>A equilibração das estruturas cognitivas</u>. São Paulo. Zahar, 1976

de maior conhecimento? Ou em outros termos, como se passa de um estado conhecimento de menor validade para um conhecimento de maior validade? [12]

De acordo com PIAGET, no processo de conhecimento, para estabelecermos relações do sujeito com o objeto, devemos rejeitar as hipóteses empiristas, pois os conhecimentos científicos, não são simples reflexos do real, mas resultam de uma atividade do sujeito que organiza e estrutura os dados da experiência a fim de compreendê-los. Opondo-se simultaneamente ao *behaviorismo* e ao *gestaltismo* e adotando uma posição dialética, o autor acima analisado afirmará:

> *"o conhecimento não procede nem da experiência única dos objetos nem de uma programação inata pré-formada no sujeito, mas de uma interação entre ambos, que resulta em construções sucessivas com elaborações constantes de estruturas novas graças a um processo de equilibrações majorantes, que corrigem e completam as formas precedentes de equilíbrio* (PIAGET, 1976:7) *".*

É preciso entender o que são essas construções sucessivas e o processo de equilibração majorante. Para esse autor, as construções sucessivas são resultados da relação sujeito-objeto na qual não há oposição, mas interação e formação de um todo único. As ações de um sobre o outro são recíprocas. Há uma perspectiva contingencial em PIAGET, para quem

> *"o sujeito constitui, com o meio, uma totalidade, sendo, portanto, passível de desequilíbrio em função das perturbações desse meio. Isso o obriga a um esforço de adaptação, de readaptação, a fim de que o equilíbrio seja reestabelecido (GIUSTA,1985: 29)"*

Para PIAGET é a ação do indivíduo, o deflagrador do desenvolvimento cognitivo e da construção do conhecimento. Na medida em que o indivíduo vai agindo no mundo, ou seja, interagindo com os objetos, ele vai transformando a realidade e os objetos (assimilação). Isso significa que vai assimilando os objetos e a realidade em suas estruturas de representação dessa mesma realidade (poderíamos dizer, do conhecimento). Tais estruturas são o conjunto articulado de elementos que o indivíduo assimilou, interligado por regras de conexões e transformações. É, portanto, uma

---

[12] Para uma discussão mais detalhada dessas questões que possuem grande complexidade e relevância ver CELLÉRIER, G. *El pensamiento de Piaget* Barcelona, Península, 1976

totalidade que se transforma e se auto-regula continuamente na medida em que o indivíduo interage com o meio. Nas palavras de PIAGET (1968):

> *"uma estrutura é um sistema de transformações que comporta leis enquanto sistema (por oposição à propriedade dos elementos) e que se conserva ou se enriquece pelo próprio jogo de suas transformações, sem que estas ultrapassem suas fronteiras ou façam apelo a elementos exteriores. Em síntese, uma estrutura compreende assim as características de totalidade, de transformações e de auto-regulação (p 6-7)".*

Mas também o indivíduo sofre a ação do seu meio, dos objetos da realidade e precisa se adaptar (acomodação) e as estruturas de representação também se adaptam. Assimilação e acomodação são processos complementares. Sem a acomodação a assimilação não seria possível. A ação do sujeito sobre o seu meio (sua interação com os objetos e a realidade através de assimilação e acomodação) promove a constante transformação das estruturas de representação. A essa contínua dinâmica de ação envolvendo assimilação e acomodação pela qual passa o indivíduo chamamos de equilibração majorante. Além da assimilação e da acomodação, é importante para viabilizar a equilibração majorante a função de organização. Analisando o processo GIUSTA (1985) afirmará:

> *"chamamos a atenção para o fato de que a assimilação/acomodação, desde os seus primórdios, apresenta-se como um ponto de partida relativo, como suporte para uma equilibração majorante, isto é para o restabelecimento do equilíbrio não apenas como uma volta ao equilíbrio anterior, mas como formação de um novo equilíbrio, ou, mais precisamente, de um melhor equilíbrio. Esse equilíbrio de nível superior funciona, então, como um novo ponto de partida, e assim sucessivamente Para que essa equilibração majorante tenha lugar, uma função de organização se faz necessária, pois, à medida que indivíduo assimila e acomoda, a organização se faz presente para integrar uma nova estrutura a uma estrutura pré-existente, que mesmo total, passa a funcionar como subestrutura. Constata-se, então, que a função de organização garante a totalidade, através da solidariedade dos mecanismos de diferenciação e de integração, preservando a continuidade e a transformação (p29)" .*

A Figura 4 abaixo pode expressar o processo de equilibração majorante: as estruturas de representação, ou seja, as estruturas de conhecimento, vão sendo reconfiguradas na medida em que o indivíduo vai interagindo com o mundo. A equilibração majorante consiste, portanto no processo de fazer com que esse conhecimento seja sempre reconfigurado a partir de constantes organizações que

ocorrem com a assimilação de novos elementos (objetos) a partir da ação do sujeito e da acomodação ou ajuste desse sujeito e de seu conhecimento ao novo, ao real.

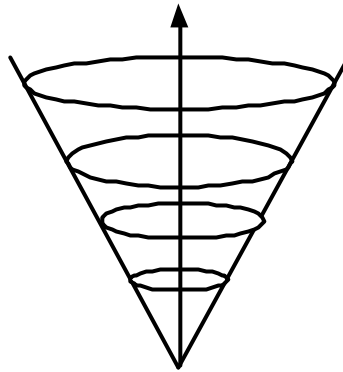



Figura 4: o processo de equilibração majorante
Fonte: Adaptado de GIUSTA (1985)

Nas palavras de PIAGET (1972):

> *"A mais bela das adaptações ao meio que a vida tenha realizado é, de fato, a adaptação do conhecimento à realidade e a mais surpreendente das evoluções, fonte de novidades e estruturas cada vez mais ricas é a evolução da própria inteligência a partir de seus começos promissores, ainda que modestos entre os animais superiores e depois se afirmando de conquista em conquista no decorrer da história humana. (...) Um primeiro resultado geral se impôs e se impõe, sempre com maior força: é o papel ativo e construtivo do sujeito pensante no ato do conhecimento, do sujeito pensante no ato do conhecimento, do sujeito pensante com relação à simples aquisição, do simples requisito das propriedades do objeto. Em outras palavras, como KANT bem vira, todo conhecimento resulta de uma síntese que depende de estruturas próprias ao sujeito, mas nós acreditamos que essas estruturas próprias ao sujeito, em lugar de serem dadas anteriormente, se constroem pouco a pouco, o que aliás reforça a parte das atividades do sujeito[13]".*

Na tentativa de tentar sintetizar a complexidade das idéias de PIAGET, podemos sugerir que o conhecimento é o estado de desenvolvimento e configuração atingido pelas estruturas de representação da realidade (e que engloba uma visão de si e uma visão do mundo), num determinado momento, fruto da organização que ocorre nessas mesmas estruturas na equilibração majorante a partir da interação do indivíduo com seu meio.

---

[13] Trecho do discurso de Jean PIAGET ao receber o Prêmio Erasmo, na Holanda em 1972 e que foi publicado na revista francesa L'Education (n. 150 – janeiro de 1973)

PIAGET insere-se, portanto, na perspectiva construtivista, um campo da teoria do conhecimento que se baseia numa concepção dialética das relações entre o sujeito que conhece e o objeto conhecido. A razão é ao mesmo tempo elemento de estruturação da realidade e estruturada por ela.

Ainda na perspectiva construtivista, merece destaque as reflexões relacionadas à "epistemologia histórica" de BACHELARD[14]. Nela temos algo análogo à idéia de equilibração majorante que é o chamado "conhecimento aproximado". Nos aproximamos da verdade a partir de sucessivos erros retificados, ou seja: o conhecimento se amplia, a verdade se manifesta e a ciência avança através de sucessivas e constantes retificações de teorias anteriores que revelam como o verdadeiro pode ser extraído do falso, numa constante polêmica em relação ao erro e num "reqüestionamento" contínuo dos saberes já objetivados. Para esse autor, o espírito científico é essencialmente retificação do saber

4.2.2.3 Pensamento, Linguagem e Conhecimento Tácito e Explícito

A questão do conhecimento não poderá ser separada do processo de linguagem e do significado que as pessoas dão aos objetos e à expressão da linguagem. POLANYI (1969) indica que as pessoas adquirem conhecimentos criando e organizando suas próprias experiências num processo interativo com a realidade. Nesse processo, é fundamental separar os conceitos de conhecimento tácito e conhecimento explícito. Em seu trabalho, *Significado e Ensino,* HAYNES (1999), analisando o trabalho de POLANYI, afirmará:

> *Michael POLANYI (1969) concorda diz que todo pensamento humano acontece dentro da existência, abarcando o significado e dominando o uso da linguagem. Para abarcar o significado temos que interiorizar as coisas externas, ou entrarmos dentro delas. O significado surge reunindo indícios em nosso próprio corpo ou fora dele. Não podemos aprender a subir numa bicicleta memorizando as leis da física que governam seu equilíbrio. As leis têm que ser feitas significativamente, tomando-se a bicicleta e tentando-se as diferentes ações que a mantêm na vertical e que a encaminham para a direção que queremos. A linguagem da física é menos necessária para a bicicleta andar do que uma compreensão do equilíbrio, a qual é aprendida fisicamente e socialmente ao mesmo tempo.*

---

[14] Gaston Bachelard (1884 – 1962).

POLANYI (1969) ilustra a necessidade e a insuficiência da linguagem para o significado com um exemplo sobre como aprendemos a descobrir uma doença pulmonar através de radiografias. À medida que o estudante presta atenção durante algumas semanas, examinando chapas de casos diferentes, um rico panorama de detalhes significantes de variações fisiológicas e mudanças patológicas, de cicatrizes e sinais de doença aguda será gradualmente revelado. As radiografias começam a fazer sentido. No mesmo momento em ele aprendeu a linguagem da radiologia pulmonar, o estudante também terá aprendido a ver significativamente as radiografias. A experiência imediata do significado da radiografia é coincidente mas não idêntica à sua consciência focal que está presente em pensamento. Conhecimento é consciência focal, aquilo de que estamos conscientes e de que podemos falar. A compreensão é a integração tácita do conhecimento. Passamos do conhecimento à compreensão sem esforço.

De acordo com POLANYI (1969), é nossa consciência auxiliar de algo que nos dota de significado. Um significado que conduz a um objeto do qual nós somos conscientes de modo focal. A compreensão tácita, que nos dá significado, consiste em coisas subsidiárias que afetam um enfoque em virtude de uma integração executada por uma pessoa.

Não podemos formular leis rígidas derivando leis gerais de experiências individuais porque cada instância de uma lei diferirá, em cada particular, de cada outra instância sua. Para formar conceitos de classe, essenciais ao significado, temos que presumir aquele processo indeterminado e global de compreensão tácita. Aplicando nossa concepção de qualquer classe de coisas, quer dores, pronomes ou pessoas, como podemos identificar objetos ou sentimentos que parecem ser diferentes em aspectos decisivos?

A compreensão tácita não pode ser reduzida à sua articulação explícita porque ela pode ser articulada por um número indeterminado de sistemas de linguagem. Quando o estudante olha uma radiografia, ele pode focá-la com a visão de um artista e vê-la em formas de luz e sombras, ou ele pode vê-la como sendo uma radiografia. Suas experiências médicas são parte da compreensão subsidiária que informará o seu foco sobre ela"[15].

A distinção estabelecida por POLANYI (1969), apresentando o conhecimento tácito como um conhecimento "interno", pessoal, próprio das pessoas e inserido nas circunstâncias, enquanto o conhecimento explícito é aquele que pode ser codificado, sistematizado e transmitido, é a base para as duas mais recentes e significativas abordagens do conhecimento organizacional: os trabalhos de NONAKA & TAKEUCHI e de DAVENPORT & PRUSAK. Num prisma organizacional, que será desenvolvido no próximo tópico, NONAKA E TAKEUCHI irão assumir as colocações de POLANYI ao afirmar que:

[15] HAYNES, Felicity Significado e Ensino. University of Wertern Austrália. Trad. Emilia Siller . A obra de Michael Polanyi referida pela autora é Knowing and Being. Chicago University Press, 1969

> *"o conhecimento que pode ser expresso em palavras e números representa apenas a ponta do iceberg do conjunto de conhecimentos como um todo. Como diz Polanyi (1969), "Podemos saber mais do que podemos dizer. Na epistemologia tradicional, o conhecimento deriva-se da separação do sujeito e do e do objeto da percepção; os seres humanos como sujeitos da percepção adquirem conhecimento mediante a análise dos objetos externos. Por outro lado, Polanyi argumenta que os seres humanos criam conhecimento envolvendo-se com objetos, ou seja, através do envolvimento e compromisso pessoal, ou o que Polanyi chama de "residir em". Saber algo é criar sua imagem ou padrão através da integração tácita de detalhes. Para entender o padrão como um todo significativo, é necessário integrar o corpo com os detalhes. Assim, o "residir em" rompe com as dicotomias tradicionais entre mente e corpo, razão e emoção, sujeito e objeto, conhecedor e conhecido. Portanto, a objetividade científica não constitui a única fonte de conhecimentos. Grande parte de nossos conhecimentos é fruto de nosso esforço voluntário de lidar com o mundo* (NONAKA E TAKEUCHI,1997: 65-66) ".

CHOO (2003) lembra que as inovações organizacionais "germinam a partir do conhecimento tácito" e que embora esse conhecimento desenvolva-se inicialmente nas habilidades e intuições dos indivíduos, tal conhecimento é gradativamente compartilhado, sendo que o grupo modula então seu conhecimento pessoal por meio de uma rede de papeis, relacionamentos e ferramentas para inventar um novo conhecimento. Portanto, o conhecimento pode emergir da articulação de pessoas e/ou organizações que compartilham informações e constroem esse conhecimento por meio de suas interações.

4.3. Prisma Organizacional

4.3.1 Uma abordagem funcional e organizacional do conhecimento

De acordo com CHOO (2003), uma organização do conhecimento é

> *"aquela que possui informações e conhecimentos que a tornam bem informada e capaz de percepção e discernimento. Num nível mais profundo, a organização do conhecimento possui informações e conhecimentos que lhe conferem uma vantagem, permitindo-lhe agir com inteligência, criatividade e, ocasionalmente, com esperteza. (...) No coração da organização do conhecimento está a administração de processos de informação, que constituem a base para criar significado, construir conhecimento e tomar decisões".*

A abordagem organizacional do estudo do conhecimento e da qual emerge mais explicitamente o termo gestão do conhecimento tem em sua gênese o reconhecimento da crescente importância do conhecimento na dinâmica organizacional. Na medida em

que a dinâmica dos mercados globalizados e os sistemas produtivos se tornam cada vez mais complexos, em especial pelos constantes investimentos em novas tecnologias e pela sofisticação das exigências para sua operação, torna-se cada vez mais necessária uma abordagem teórica e organizacional que responda a esses desafios. Essa nova abordagem pode passar pelo conceito de gerência do conhecimento. De certa maneira, os principais conceitos que são apresentados pelos principais autores dessa abordagem, tem fortes raízes na abordagem filosófico-cognitiva apresentada na seção anterior. Entretanto há uma perspectiva funcionalista no uso dos conceitos em favor da otimização da dinâmica da organização, o que leva muitas vezes interpretações simplistas e que muitas vezes não explicitam as contribuições recebidas. Na definição de NONAKA e TAKEUCHI (1997)

> "*criação do conhecimento organizacional é a capacidade que uma empresa tem de criar conhecimento, disseminá-lo na organização e incorporá-lo a produtos, serviços e sistemas e o sucesso alcançado pelas empresas japonesas se deveu em grande parte à sua capacidade e habilidade em criar e aplicar o conhecimento organizacional* (NONAKA E TAKEUCHI,1997: Cap 1) ".

Conforme já mencionado, o trabalho de NONAKA e TAKEUCHI pode ser considerado um marco no estudo recente do que podemos chamar de visão organizacional do conhecimento. Em sua teoria, eles estão a propor um modelo de "espiral do conhecimento"[16] baseado no confronto de três dimensões, que classificam de ontológica e epistemológica, conforme apresentada na Figura 5 abaixo. A terceira dimensão é o tempo, que é expresso por um conjunto de fases no processo de criação do conhecimento organizacional.

[16] No Capítulo 3 NONAKA E TAKEUCHI (1997) descrevem de modo sistematizado o que chamam de Teoria da Criação do Conhecimento Organizacional, cuja essência apresentamos no resumo que se segue a esta descrição. Conceitos filosóficos são resgatados pelos autores ao longo da obra e colocados numa perspectiva funcionalista para a organização contemporânea.

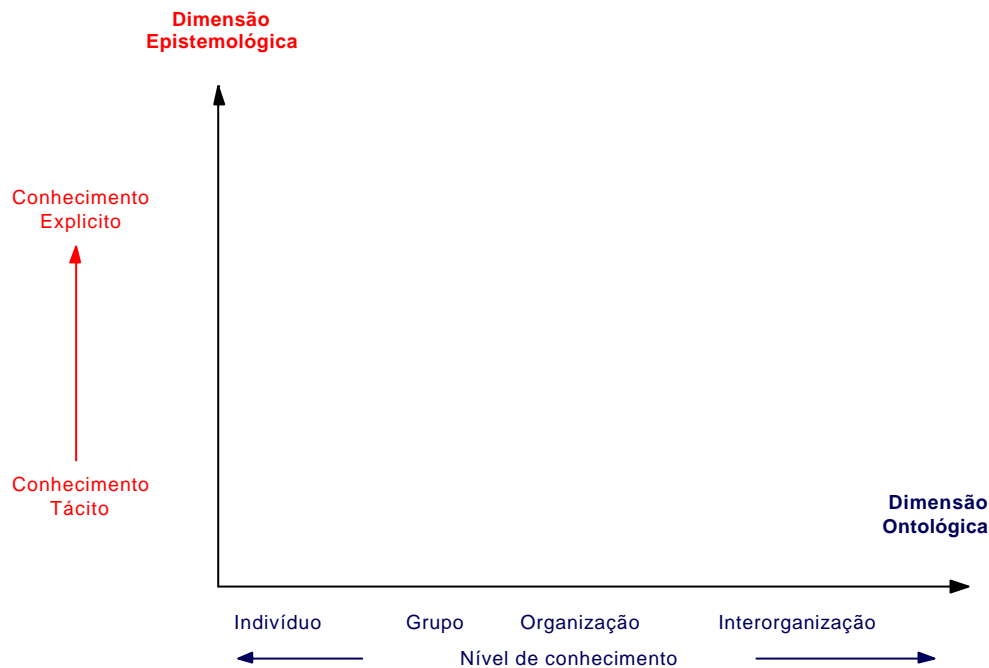



Figura 5: Duas dimensões de criação do conhecimento
Fonte: NONAKA E TAKEUCHI (1997: 62)

Em termos ontológicos, apresentam a visão de que o conhecimento numa organização não pode ser criado sem indivíduos. Na perspectiva de STEWART (1998) esses indivíduos podem ser os empregados, os clientes, os fornecedores ou os parceiros e todos precisam ser ouvidos. O papel da organização é estimular os indivíduos criativos e oferecer-lhes contexto para que criem conhecimento. Apreciar suas idéias e valoriza-las, estimulando ambientes que favoreçam a partilha de experiências e a criação do conhecimento. Esse conhecimento pode ser expandido "organizacionalmente", cristalizando-se como parte da rede de conhecimentos da organização dentro de uma "comunidade de interação". Uma expansão que poderia, inclusive, extrapolar as fronteiras organizacionais. Essa análise pode salientar o papel importante da cultura organizacional e do trabalho coletivo. SHEIN (1986), citado na obra de FLEURY (1997), afirma que

> *"a cultura é formada pelo conjunto de pressupostos básicos que um grupo inventou, descobriu, ou desenvolveu, ao aprender a lidar com os problemas de adaptação externa e integração interna e que funcionaram bem o suficiente*

para serem considerados válidos e ensinados a novos membros como a forma correta de perceber, pensar e sentir com relação a esses problemas”[17].

Na mesma obra de FLEURY (1997) é exposto um enriquecimento deste conceito:

“*cultura organizacional é um conjunto de valores, expressos em elementos simbólicos e em práticas organizacionais, que em sua capacidade de ordenar, atribuir significações, construir a identidade organizacional, tanto agem como elementos de comunicação e consenso como expressam e instrumentalizam relações de dominação*”[18].

Conjunto de valores e práticas que fazem com que a organização possa agir no seu ambiente expressam simultaneamente a cultura e o conhecimento organizacional. Realmente são as pessoas que criam o conhecimento, mas quando essa criação se dá dentro de um ambiente e dentro de determinadas circunstâncias podemos ter o conhecimento organizacional, ou seja: um conhecimento que foi gerado pela interação de pessoas atuando sob um conjunto de princípios, valores, objetivos e estratégias. Além disso, esse conhecimento só faz seu maior sentido dentro daquele ambiente, isto é: os resultados ou transformações que podem produzir estão em conexão com esse contexto organizacional. Abre-se, portanto, uma perspectiva sistêmica para a questão do conhecimento.

Em seu estudo sobre sistemas, BERTALANFFY (1968), postula que as propriedades dos sistemas não podem ser descritas significativamente em termos de seus elementos separados e sua compreensão irá ocorrer apenas quando estudarmos os sistemas globalmente[19]. CHIAVENATO (1976) afirma que “o holismo é a tese que sustenta que algumas totalidades representam mais do que a soma das partes” e, citando o trabalho de *Jan Christhian* SMUTS (1926), *Holismo e Evolução,* descreve que “ao serem reunidos para constituir uma unidade funcional maior, os componentes individuais de um sistema desenvolvem qualidades que não são produzidas a partir de seus comportamentos isolados”.[20] Portanto, numa perspectiva organizacional e

---

[17] SHEIN, E. Organizational culture and leadership. San Francisco: Jossey-Bass, 1.ed. 1986, *in* FLEURY, A . & FLEURY, M. Aprendizagem e Inovação Organizacional: as experiências de Japão, Coréia e Brasil São Paulo: Atlas, 1997. P. 24

[18] FLEURY, M.T. Cultura e poder nas organizações. São Paulo. Atlas, 1989

[19] BERTALANFFY, Ludvig von. *General Systems Theory*. Nova York, George Brasilier, 1968

[20] CHIAVENATO, Idalberto. Introdução à Teoria Geral da Administração. 5ª. ed. São Paulo, Makron Books, 1976

sistêmica, que focaliza mais a interação entre elementos componentes de um sistema, no caso o sistema seria a organização, do que os elementos em si, no caso os elementos são as pessoas, poderíamos afirmar que o conhecimento organizacional emerge do inter-relacionamento entre as pessoas que trabalham juntas no seio de uma cultura, permanecendo "vivo" ou existente dentro dessa mesma cultura ou quando será formalizado ou explicitado em normas, regras, planos, documentos e sistemas de informação.

Dentro dessa perspectiva temos dois conhecimentos que se relacionam: um conhecimento informal, fluído nas habilidades das pessoas e na sua interação, mas que gradativamente pode, ao menos em parte, ir sendo explicitado formalmente em sistemas de controle e gestão das organizações.

Essa análise se coaduna com a visão de NONAKA E TAKEUCHI (1997) que podemos retomar, agora em termos epistemológicos. Na opinião desses autores, as empresas precisam detectar e valorizar o conhecimento tácito, ou seja, o que sabemos implicitamente, por dentro, e o que sabemos formalmente. O conhecimento tácito é pessoal, especifico do contexto e, portanto, difícil de formalizar e comunicar. O conhecimento explícito ou "codificado", por outro lado, refere-se ao conhecimento que é transmissível na linguagem formal e sistematizada. Por isso, será de fundamental importância a conversão do conhecimento tácito em explicito, processo no qual os gerentes de nível médio vão desempenhar importante papel, na medida em que captar o conhecimento tácito dos empregados dos níveis operacionais e também daqueles da alta direção, tornando-o explicito e incorporando-o a novas tecnologias. Na viabilização da utilização do conhecimento tem-se um processo de conversão do conhecimento, caracterizado por quatro modos que se interagem de modo cíclico no processo de conversão do conhecimento. Esses quatro modos são:

- Socialização: descreve o compartilhamento implícito de conhecimento tácito, muitas vezes sem uso de linguagem, como por exemplo através da experiência, que pode ser passada a outra pessoa até mesmo pela observação, imitação e prática. NONAKA & TAKEUCHI afirmam que se trata de "um processo de compartilhamento de experiências e, a partir daí, da criação do conhecimento tácito, como modelos

mentais ou habilidades técnicas compartilhadas” Uma nota no trabalho desses autores esclarece o conceito de modelo mental, na qual se afirma

> *“CANNON-BROWERS, J.A ., E. SALAS, e S. CONVERSE (1993) definem modelos mentais compartilhados como sendo estruturas de conhecimento mantidas pelos membros de uma equipe que lhes permitem elaborar explicações e expectativas precisas para a tarefa e, em contrapartida, coordenar suas ações e adaptar seu comportamento à demandas da tarefa e a outros membros da equipe”.(...) Para compreender como um modelo mental compartilhado é criado, é útil conhecer o conceito de “fusão de horizontes” do filósofo alemão Hans-Georg GADAMER (1989) que argumenta que uma verdadeira compreensão de um texto é uma “fusão dos horizontes do intérprete e do autor. Define o horizonte como “a faixa de visão que inclui tudo o que pode ser visto de um determinado ponto de observação (p.302). Aplicando esse conceito ao nosso conceito, podemos argumentar que a socialização é uma “fusão” dos conhecimentos tácitos dos participantes em um modelo mental compartilhado”*[21].

- Externalização: significa converter o conhecimento tácito em conhecimento explícito. Isso pode se dar através de métodos analíticos, como dedução e indução através da formalização objetiva de idéias, obtida pela linguagem na criação de conceitos, hipóteses e modelos. Mas, quando a linguagem não é competente o suficiente para externalizar idéias, ainda assim há os recursos da utilização de metáforas, analogias. “O modo de externalização da conversão do conhecimento normalmente é visto no processo de criação do conceito e é provocado pelo diálogo ou pela reflexão coletiva”, afirmam NONAKA E TAKEUCHI (1997). Além disso, “é fundamental considerar que externalização é a chave para a criação do conhecimento, pois cria conceitos novos e explícitos a partir do conhecimento tácito”.[22]
- Combinação: conforme os atores do modelo, “é um processo de sistematização de conceitos em um sistema de conhecimento. A reconfiguração das informações existentes através da classificação, do acréscimo, da combinação e da categorização do conhecimento explicito (como o utilizados em bancos de dados de computadores) pode levar a novos conhecimentos”. Bancos de dados, redes de informação

[21] Análise realizada por NONAKA & TAKEUCHI, opus cit. P. 69 e p. 297. As obras citadas pelos autores são: CANNON-BROWERS, J.A ., E. SALAS, e S. CONVERSE. Shared Mental Models in Expert Team Decision Making, org. N.J.CASTELAN, JR., pp 221-246. HILLSDALE, NJ: Lawrence Erbaum Associates. 1993. Também: GADAMER, H. Truth and Method. 2ª. ed. para o inglês de J. Weinsheimer e D.G. Marshal. Nova York. Crossroad. 1989.

[22] Para uma anális mais detalhada, conferir em (NONAKA E TAKEUCHI,1997: 71-73).

computadorizadas, sistemas de informação e cursos, como de pós-graduação são bons exemplos de elementos que viabilizam a combinação de conhecimentos.

- Internalização: consiste na "incorporação do conhecimento explícito no conhecimento tácito" As pessoas que trabalham, internalizam em suas bases de conhecimentos tácitos as experiências e oportunidades de aprendizado que tiveram. Isso é feito sob a forma de construção de novos modelos mentais e absorção de know-how técnico, os quais poderão ser utilizados no futuro. Esse processo é auxiliado por processos de verbalização e diagramação do conhecimento, através de descrições coletivas de histórias orais, documentações e manuais elaborados para promover uma nova experiência no futuro junto a outras pessoas.[23]

A partir do exposto acima, NONAKA & TAKEUCHI propõe o modelo chamado de "espiral do conhecimento" [24]. Não se tornando explicito, o conhecimento compartilhado não pode ser facilmente utilizado pela organização como um todo. O processo de inovação[25], por exemplo, surge da interação dinâmica entre o conhecimento tácito e conhecimento explicito, o que significa que experiências são socializadas e compartilhadas até que possam ser externalizadas pelo diálogo e pela reflexão coletiva, empregando-se para isso recursos analíticos, metáforas ou analogias. Em seguida, esse conhecimento novo pode ser confrontado com aquele já existente num processo de combinação. Por último, um processo de "aprender fazendo"provoca a inovação. A abordagem desses autores pode ser representada no diagrama seguinte, no qual também são explicitados o conteúdo do conhecimento criado por cada modo de conversão do conhecimento. Os conteúdos do conhecimento são:

- Conhecimento compartilhado (habilidades técnicas e modelos mentais, por exemplo);
- Conhecimento conceitual, fruto da externalização por metáforas e processos analíticos de reflexão, conhecimento sistêmico (caracterizado pela combinação de diversos elementos que podem gerar, por exemplo, novas tecnologias e produtos);

[23] Ibid p. 77-78
[24] Ibid p. 79-82
[25] A inovação é uma expressão relevante da geração e aplicação concreta do conhecimento dentro de uma organização, bem como de sua relevância:

- Conhecimento operacional (expresso pela capacidade de internalizar o uso de habilidades e executar tarefas e processos diversos).

Figura 6: Espiral do Conhecimento
Fonte: (NONAKA E TAKEUCHI,1997: 80-81)

Continuando a exposição da Teoria da Criação do Conhecimento Organizacional de NONAKA & TAKEUCHI, devemos lembrar que esses conteúdos do conhecimento interagem continuamente entre si numa outra espiral, a da criação do conhecimento, que enfoca o que eles caracterizam como sendo a "*dimensão ontológica*" do modelo. Conforme afirmam:

> *"O conhecimento tácito dos indivíduos constitui a base da criação do conhecimento organizacional. A organização tem de mobilizar o conhecimento tácito criado e acumulado no nível individual. O conhecimento tácito mobilizado é ampliado "organizacionalmente" através dos quatro modos de conversão do conhecimento e cristalizado em níveis ontológicos superiores. Chamamos isso de "espiral de criação do conhecimento", na qual a interação entre conhecimento tácito e conhecimento explícito terá uma escala cada vez maior na medida em que subirem os níveis ontológicos. Assim, a criação do conhecimento organizacional é um processo em espiral, que começa no nível individual, e vai subindo, ampliando comunidades de interação que cruzam fronteiras entre seções, departamentos divisões e organizações. Um bom exemplo desse processo é exemplificado pelo desenvolvimento de produtos* (...) *A função da organização no processo de criação do conhecimento organizacional é fornecer o contexto apropriado para facilitação das*

atividades em grupo e para criação e acúmulo de conhecimento em nível individual"[26].

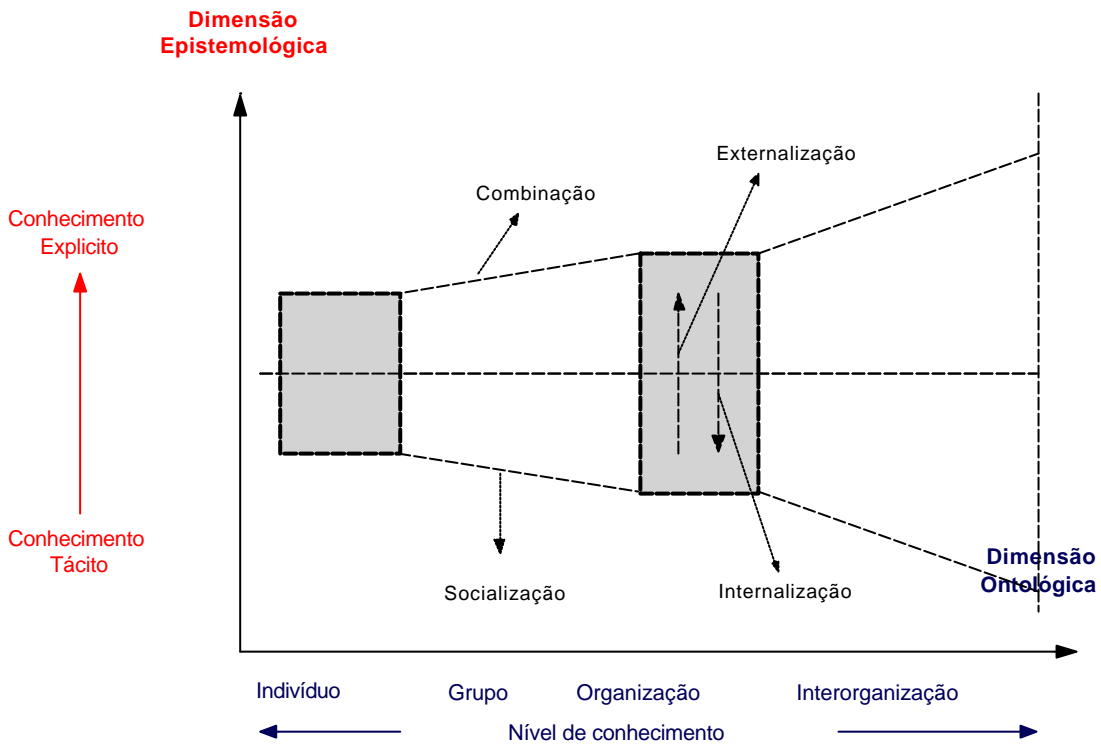
A espiral de criação do conhecimento organizacional é apresentada na Figura 7 abaixo:

Espiral de criação do conhecimento (NONAKA E TAKEUCHI, 1997)

Figura 7: Espiral de criação do conhecimento organizacional
Fonte: (NONAKA E TAKEUCHI,1997: 82)

Incorporando a dimensão tempo ao modelo de NONAKA & TAKEUCHI[27], teremos a visão de que o processo de criação do conhecimento organizacional ocorre numa seqüência de cinco fases, que são:

- Compartilhamento do Conhecimento Tácito: É o primeiro desafio do processo de criação do conhecimento organizacional e significa que os indivíduos devem ter a possibilidade de interagir uns com os outros levando em conta suas múltiplas diferenças. Essa é uma etapa crítica, pois como afirmam esses autores, "uma vez que o conhecimento adquirido com a experiência nem sempre é facilmente articulável pela linguagem formal. Além disso, emoções, sentimentos e modelos mentais tem de

---

[26] Ibid. p 83
[27] Ibid. Conferir análise p. 95-102

ser compartilhados para permitir o desenvolvimento da confiança mútua" Equipes auto-organizadas, nas quais pessoas de área diferentes trabalham em conjunto para alcançar uma meta comum é um campo excelente para viabilizar esta fase.

- Criação de Conceitos: É a fase que corresponde à externalização, já citada. Implica que as pessoas conseguem formalizar, explicitar o conhecimento. "A criação de conceitos pode se dar de forma cooperativa, através do diálogo e para isso a autonomia é importante.
- Justificação de Conceitos: A justificação significa a determinação de que conceitos novos que foram desenvolvidos tem algum valor para a organização ou a sociedade. Basicamente é um processo de filtragem e validação do novo, do recém criado à luz de critérios que sejam considerados como relevantes, tais como: custo, margem de lucro, grau de inovação, relevância para uma empresa. Tais critérios podem ser tanto quantitativos quanto qualitativos.
- Construção de Arquétipos: É a concretização de uma idéia ou conhecimento que foi validado e aceito. Um conceito que foi justificado precisa ser transformado em algo tangível ou concreto, que é o arquétipo, ou modelo. Esse processo coincide com a combinação, na qual o conhecimento novo se funde com o antigo para produzir resultados reais. É como um protótipo, uma "maquete" desenvolvida por um arquiteto para expressar seu projeto.
- Difusão Interativa do Conhecimento: Expressa uma realidade dinâmica na qual a geração de um conhecimento concretizado reinicia o ciclo gerando ainda mais novos conhecimentos. Um modelo de produto, uma inovação em um processo continua e reinicia novamente o ciclo de criação do conhecimento num processo contínuo de inovação.

Outro trabalho igualmente importante relativo à questão da gestão do conhecimento é o de DAVENPORT & PRUSAK (1998). Esses autores afirmam que o conhecimento é um ativo corporativo, constituindo a maior vantagem competitiva da empresa e que, por isso precisa ser gerido com o mesmo cuidado dedicado a outros ativos tangíveis. Avaliam que as vantagens competitivas das empresas possuem ciclos cada vez mais curtos e que o conhecimento é uma vantagem com maior sustentabilidade, uma vez que:

> *"Com o tempo, os concorrentes quase sempre conseguem igualar a qualidade e o preço do atual produto ou serviço do líder do mercado. Quando isso acontece, porém a empresa rica em conhecimento e gestora do conhecimento terá passado para um novo nível de qualidade, criatividade ou eficiência. A vantagem do conhecimento é sustentável porque gera retornos crescentes e dianteiras continuadas* (DAVENPORT & PRUSAK, 1998:14-20)".

Questões centrais são a idéia de que o porte de uma organização e sua dispersão geográfica tornam difícil o gerenciamento do "estoque de conhecimento" de uma empresa globalizada e a visão de que os computadores, cada vez, mais acessíveis em termos de custo e "operabilidade" são ferramentas cada vez mais poderosas como propiciadores de conhecimento, na medida em que possuem recursos de armazenagem e comunicação de dados e produção de informações na medida em que redes dessas máquinas podem integrar as pessoas.

DAVENPORT & PRUSAK (1998) procedem a uma análise comparativa do conhecimento como se houvesse juntado à organização um "mercado". Nesse "mercado do conhecimento" afirmam existir *compradores* (pessoas que estão tentando resolver algum problema cuja complexidade ou incerteza não permite uma resposta fácil), *vendedores* (aqueles que desenvolvem ou possuem algum conhecimento relevante) e *corretores* (todo profissional que coloca as duas figuras anteriores em contato (podem ser bibliotecários, pesquisadores, analistas de informação). A relação entre os agentes possui um "sistema de preços" específico que inclui reciprocidade, reputação, altruísmo (quem oferece conhecimento um dia precisará requisitar, em geral, deseja ser reconhecido explicitamente, embora possa algumas vezes ajudar sem receber nada em troca). Isso só pode ser obtido num ambiente de grande confiança entre esses agentes. Cabe à organização evitar "ineficiências e patologias" do mercado de conhecimento, tais como os monopólios (apenas alguns detém conhecimento) e as barreiras diante de inovações de idéias e práticas organizacionais (DAVENPORT & PRUSAK, 1998: 36-53)." Afirmando que

> "*organizações geram e usam conhecimento. à medida que interagem com seus ambientes, elas absorvem informações, transformam-nas em conhecimento e agem com base numa combinação desse conhecimento com suas experiências, valores e regras internas* (DAVENPORT & PRUSAK, 1998: 63)"

esses autores descrevem o que classificam como os cinco modos de geração consciente e intencional de conhecimento, isto é, iniciativas específica que as organizações empreendem para aumentar o estoque de conhecimento corporativo. Tais modos são:

- Aquisição: o mais importante para a organização é a utilidade e não a originalidade do conhecimento. Portanto, a organização tem que contar com a disponibilidade do conhecimento apropriado quando e onde ele possa ser aplicado. Por isso, muitas vezes a melhor e mais rápida maneira de obter um conhecimento necessário poderá ser a aquisição de indivíduos ou pessoas que o possuem. Muitas vezes pode significar a compra de empresas inteiras que detém algum tipo relevante e específico de conhecimento ou competência.
- Aluguel: é uma variação do processo de aquisição e significa que uma empresa pode terceirizar parte do desenvolvimento de seus produtos e serviços ou solução de sérios problemas. Pode implicar desde parcerias acadêmicas até contratos com consultores.
- Recursos Dirigidos: significa que a organização destina recursos diversos, como departamentos de pesquisa e desenvolvimento, bibliotecas corporativas, centros de informação, softwares e etc para viabilizar o desenvolvimento do conhecimento que interessa a ela.
- Fusão: significa estimular a geração do conhecimento pelo confronto de diferentes perspectivas, opiniões e pontos de vista, mesmo que inicialmente seja necessário introduzir alguma desordem ou "caos criativo". Isso parte da perspectiva de que inovações em geral ocorrem justamente do confronto entre diferentes pontos de vista e da combinação de diferentes opiniões, trabalhos e habilidades.
- Redes: grupos formais ou informais, comunidades *"ad hoc"* ou auto-organizadas também são responsáveis pela geração do conhecimento. As pessoas se aglutinam, motivadas por interesses comuns, e em geral, podem conversar pessoalmente, por telefone, correio eletrônico ou *groupware* (DAVENPORT & PRUSAK, 1998: 64-81)."

4.3.2 Codificação, Taxonomia, Aplicabilidade e Transferência do Conhecimento

4.3.2.1 Codificação e Taxonomia do Conhecimento

DAVENPORT & PRUSAK (1998) apresentam uma análise sobre a codificação e criação do conhecimento, uma vez é necessário apresentar o conhecimento de uma forma acessível a todos os públicos de uma organização, embora lembrem que nem todo conhecimento pode ser codificado. Em sua opinião, os gerentes devem decidir a que objetivos o conhecimento codificado irá servir, em quais formas ele existe dentro das organizações, qual sua utilidade e quais os meios apropriados para codificação e disseminação. Essa perspectiva é realmente importante, pois nem todo conhecimento interessa a uma organização, o que implica que sempre haverá algum juízo de valor na determinação daquilo que interessa ou não a ela. Esses autores afirmam que o conhecimento tácito é complexo e é desenvolvido e interiorizado pelo conhecedor no decorrer de um longo período de tempo. É quase impossível de reproduzi-lo num banco de dados ou documento. DAVENPORT & PRUSAK (1998) descrevem um quadro que apresenta um "*spectro*" das dimensões de codificação do conhecimento conforme descrito abaixo:

Tácito ______________________________ Articulável
Não passível de ensino ______________________________ Passível de ensino
Não Articulado ______________________________ Articulável
Não observável em uso ______________________________ Observável em uso
Substancial ______________________________ Esquemático
Complexo ______________________________ Simples
Não documentável ______________________________ Documentado

É importante salientar uma questão central abordada por DAVENPORT & PRUSAK (1998): a taxonomia do conhecimento. A separação e o reconhecimento da existência dos conhecimentos tácito e explicito, já citados neste trabalho, são retomados e apresenta-se um imperativo de se "codificar o conhecimento" (DAVENPORT & PRUSAK, 1998: 85-105)." Realmente, para efeito do processo gerencial de uma organização e importante voltar atenção para o conhecimento que possa ser, de alguma forma codificado ou sistematizado e para aquele conhecimento que, embora tácito, pode

ser explicitado. Ou então, é importante considerar o conhecimento tácito em sua própria natureza, mas é fundamental que ele possa ser, pelo menos, localizado em “mapas do conhecimento”, isto é, processos e intervenções gerenciais que conduzam a criação, manutenção e uso do conhecimento na organização. Uma atenção ainda mais especial pode ser dada para o conhecimento explicito, uma vez que pode ser apoiado pela tecnologia da informação. Há uma percepção muito importante no trabalho dos últimos autores citados ao afirmarem:

> *“o conhecimento tácito e complexo, desenvolvido e interiorizado pelo conhecedor no decorrer de um longo período de tempo é quase impossível de se reproduzir num documento ou banco de dados (...) Simplesmente alguns tipos de conhecimento não podem ser representados efetivamente fora da mente humana e é por isso que o processo de codificação do mais substancial conhecimento tácito existente nas organizações se limita a localizar alguém que possua aquele conhecimento, encaminhar o interessado para aquela pessoa e estimular ambos a interagir*(DAVENPORT & PRUSAK, 1998: 86-87).”

Sobre a montagem dos “mapas de conhecimento organizacional” irão afirmar que se monta um mapa público a partir de mapas individuais :

> *“As informações necessárias para se criar um mapa do conhecimento normalmente já existem nas organizações, porém de forma fragmentada e não documentada. Cada pessoa tem um pedacinho do mapa na cabeça, sabe qual é o seu próprio conhecimento e onde ir para obter respostas para as suas perguntas. Criar um mapa organizacional é uma questão de combinar esses mini-mapas individuais em um mapa corporativo”* (DAVENPORT & PRUSAK, 1998: 90)”.

Na opinião desses autores, a organização pode pesquisar junto aos funcionários que conhecimentos eles possuem, suas fontes e em que esse conhecimento é útil. Pode também se servir de um processo de trilha de recomendações, isto é: uma pessoa sugere outras que conhecem assunto as quais por sua vez vão, em cadeia, sugerindo outras, até que a organização delimite o conhecimento desejado.

4.3.2.2 Aplicabilidade e Transferência do Conhecimento

O conhecimento ou um corpo de conhecimentos pode ser classificado de diferentes formas. A classificação por domínio, ou seja, do que trata um determinado conhecimento, quais assuntos e temas estão embutidos nesse domínio é muito comum.

A partir de estudos no *Ernst & Young Center of Business Innovation*, NOVINS & ARMSTRONG (2002) afirmam que existem maneiras de classificar o conhecimento que tem pouca relação com o conteúdo. Ao invés de se indexar um conhecimento tendo como base se ele "é sobre isto ou aquilo", por exemplo, nós podemos classificá-lo com base em sua origem (de onde ele pode ser obtido). Ele provém de uma única fonte, ou de múltiplas fontes? Esta é uma classificação com foco na origem, e não no domínio como é mais comum se encontrar. Outras formas de classificação do conhecimento incluem:

- Pelo receptor: quem irá necessitar utilizá-lo?
- Pela aplicabilidade: Qual a amplitude da utilização do conhecimento? Ele é de natureza local ou global?
- Pela transferência: Qual o grau de dificuldade em realizar sua transferência aos outros, e qual o grau de dificuldade para os outros ao aplicar corretamente o conhecimento recebido?
- Pela qualidade: quanto um determinado conhecimento é dependente de seu contexto, e quanto será perdido em sentido através de representações, explicitações ou simplificações?
- Pela perenidade: qual a idade do conhecimento? O conhecimento continua válido independentemente da passagem do tempo?
- Pela confiabilidade: È fácil de ser conferido (testado)? Provém de fonte confiável?

O conjunto de critérios poderia continuar de forma extensiva. O ponto é que, ao considerar a questão "*qual conhecimento compartilhar*", a resposta não precisa ser obrigatoriamente expressa como conhecimento sobre x ou sobre y. Ao avaliar quais conhecimentos devemos gerenciar e como devemos fazer isso, pode ser mais útil, dependendo da perspectiva, considerar sua aplicabilidade e transferência, do que raciocinar em termos de domínio.

De fato, entendemos que pensar no conhecimento exclusivamente em termos de domínio às vezes pode não ser muito útil em se tratando de orientar o seu gerenciamento. Uma nova visão surge ao analisarmos o conhecimento quanto a seu

grau relativo de aplicabilidade e transferência. NONAKA E TAKEUCHI (1997) trataram dessa questão sob uma determinada perspectiva que agora podemos complementar e que implica na idéia de que todo o conhecimento que pode ser importante para as decisões de um negócio pode ser classificado numa linha crescente de aplicabilidade e transferência, ou seja: "*plotado*", inserido em ponto ao longo de um "*continuum*" quanto à possibilidade de aplicação e transferência. Tais idéias são elementos básicos da visão de NOVINS & ARMSTRONG (2002), sintetizada na Figura 8, que é uma matriz com quatro categorias de conhecimento possíveis, cujos títulos apontam para diferenças chave na forma como eles devem ser gerenciados. Nessa matriz, a aplicabilidade se encontra no eixo horizontal e a transferência no eixo vertical. Segundo tais autores na linha horizontal o conhecimento é classificado sob o ponto de vista da aplicabilidade, variando em um "*continuum*" de local para global. Em um extremo, um conhecimento sobre alguma coisa pode ser de natureza puramente local. Isto quer dizer que ele se aplica exclusivamente a um conjunto especifico de condições, sendo dependente de uma dada situação geográfica ou física. Em outro extremo, o conhecimento possui natureza global, podendo ser aplicado largamente através de toda uma organização, e através de fronteiras de processos, técnicas, e culturais. Uma maneira de se raciocinar sobre esta forma de classificação é pensando em termos de "conhecimento detalhado" versus "conhecimento geral". Infere-se então que um dado conhecimento pode ser gerenciado diferentemente dependendo da amplitude de sua aplicação. Já na linha vertical, o conhecimento é classificado entre único e programável, em que é considerado então o grau de transferência exibido por um certo conjunto de conhecimentos Conhecimento que pode ser classificado com base em regras é altamente transferível, pois ele pode ser facilmente recuperado a qualquer tempo. (se a condição x está presente, então a melhor abordagem é y). Ele pode ser transferido múltiplas vezes sem perder sua validade. No outro extremo, a transferência é baixa Olhar o conhecimento em termos de aplicabilidade e transferência nos dá uma clara direção de como armazenar e transferir conhecimento. A Figura 8, portanto, resume os quatro "*tipos*" de conhecimento dentro da visão de NOVINS & ARMSTRONG (2002). Para cada um dos quatro tipos de conhecimento mostrados, é definida uma forma de gerenciamento, como segue:

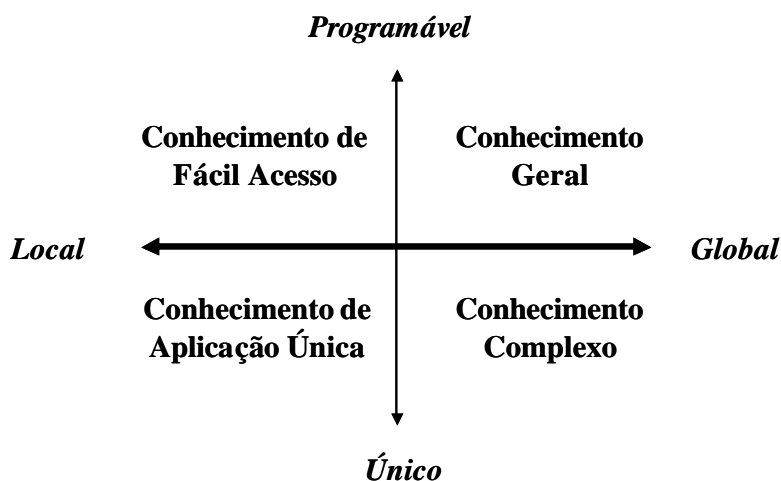



Figura 8 – Quatro categorias do conhecimento .
Fonte: NOVINS & ARMSTRONG (2002) p. 48

- *Conhecimento de Acesso Rápido:* È gerenciado através de sua colocação em um local facilmente acessível, na maioria das vezes através de sofisticados bancos de dados. Seria um erro distribuir este conhecimento pró-ativamente para todo o pessoal, só pelo fato de alguém, algum dia, poder necessitar dele para tomar uma decisão específica. Um exemplo deste tipo de conhecimento são relatórios técnicos emitidos, por institutos de pesquisa, cuja existência seria informada a poucas pessoas, e ficariam em lugar de fácil acesso.

- Conhecimento Geral*:* Estes conhecimentos são de fácil acesso e de uso geral dentro da empresa. Vale a pena, portanto, divulgá-los *pró-ativamente* através de circulação e publicação na *Intranet*. Um exemplo é uma instrução que regulamente viagens a serviço.

- Conhecimento Complexo: São conhecimentos de difícil acesso, via de regra com grande parcela de conhecimento tácito, e que são de interesse de um grande número de pessoas na empresa. Este conhecimento é melhor transferido através de programas de treinamento estruturados e através do aprendizado no trabalho, utilizando-se, por exemplo o "mentoring". Um exemplo deste caso é o gerenciamento de um empreendimento.

- Conhecimento de Aplicação Única: São conhecimentos difíceis de serem transferidos, tem grande parcela de conhecimento tácito, e de aplicação restrita. Não se deve investir muito na gestão deste tipo de conhecimento, a não ser no caso de algo de excepcional relevância. estratégica.

Sob a perspectiva acima analisada, sugere-se que pensar sobre conhecimento em termos de aplicabilidade e transferência resulta numa orientação bem mais efetiva para o seu gerenciamento do que se pensarmos em termos de domínio. Nos quatro casos, assim como em muitas áreas de negócio, a melhor forma de gerenciamento é um balanço cuidadoso entre influenciar o comportamento das pessoas, introduzir processos efetivos, e disponibilizar tecnologia de suporte. O "*mix*" vai diferir, entretanto, com a categoria. Conhecimento de acesso rápido, por exemplo, é forte candidato à "informatização", e o gerenciamento em tal caso deverá ser intensivo em tecnologia da informação. Conhecimento complexo, por outro lado, demanda o mais alto grau de gerenciamento de pessoas. As quatro categorias comportam claras implicações quanto ao nível de investimento e esforços necessários para o seu gerenciamento. Conhecimento de aplicação única pode oferecer baixo retorno para um alto esforço de gerenciamento. O gerenciamento do conhecimento complexo pode representar grande fonte de vantagem competitiva.

NOVINS & ARMSTRONG (2002) também fazem um análise do conhecimento quanto a sua forma de compartilhamento, a qual fornece importantes subsídios para a melhor forma de geri-lo. A figura 9 mostra diferentes possibilidades de compartilhamento, e apresenta mecanismos para efetivá-lo. Uma pergunta é colocada: "*Entre quem o conhecimento precisa ser compartilhado?*" e tal pergunta pode ser respondida sob o ponto de vista qualitativo (quem são as pessoas) e quantitativo (quantas pessoas estão envolvidas). O ponto de vista quantitativo coloca desafios importantes e precisa ser bem visualizado, e isso significa que o conhecimento pode ser compartilhado de:

- um-para-um;
- um-para-muitos;
- muitos-para-um;

- muitos-para-muitos.

Os mecanismos para efetivar o compartilhamento do conhecimento propostos são:

- De um-para-um – aprendizagem, *mentoring* e *coaching.*[28]
- De um-para-muitos – apresentações, livros e artigos.
- De muitos-para-um – redes.
- De muitos-para-muitos – alavancagem.[29]

Os indivíduos sentem-se mais confortáveis com o compartilhamento de conhecimento que provém de outros indivíduos. Uma pessoa detentora de conhecimento se imagina transferindo conhecimento para outra pessoa, ou para muitas outras pessoas. Similarmente, ao se imaginar recebendo novos conhecimentos, pensa em recebê-los de alguma outra pessoa que está se dirigindo a ela diretamente ou a considerando como membro de um grupo.

---

[28] *Mentoring* é uma expressão utilizada no campo da Administração de Recursos Humanos – ARH que significa um processo de orientação direcionada de carreira, envolvendo aconselhamento, identificação e estimulo de habilidades e competências com planejamento das estratégias de desenvolvimento das mesmas ao longo do tempo. Já a palavra *coaching* é utilizada em ARH no sentido de formação e capacitação profissional direcionada, envolvendo atividades de educação profissional e treinamento.

[29] O termo alavancagem – resultado da força de uma alavanca – indica aqui o resultado sinérgico e vigoroso obtido pela mesclagem e compartilhamento de múltiplos conhecimentos de diferentes pessoas na organização. Conforme NOVINS & ARMSTRONG (2002), a transferência de conhecimento na modalidade "muitos para muitos" certamente não vai se tornar confortável da noite para o dia. Os gerentes nas organizações irão continuar com a tendência de aprender da forma tradicional, individualmente, e continuarão acreditando em indivíduos. Ao se depararem com uma troca muito mais ampla irão impor resistência por experimentarem uma sensação de perda de controle. (é muito conhecimento sendo distribuído de uma forma muito ampla). No entanto, apesar destas dificuldades, cada vez mais está se reconhecendo que a real alavancagem obtida dos ativos do conhecimento é alcançada com esta modalidade de compartilhamento, e que o gerenciamento do conhecimento deve focar nela primeiro.

	UM	MUITOS
UM	APRENDISAGEM MENTORING COACHING	REDES
MUITOS	APRESENTAÇÕES LIVROS ARTIGOS	ALAVANCAGEM

Figura 9 – Diferentes Níveis de Compartilhamento do Conhecimento
Fonte: NOVINS & ARMSTRONG (2002) p. 51

Uma grande organização pode "aprender" mais em um dia do que uma pessoa em toda a sua carreira. Desta forma faz sentido, que quando alguém necessitar tomar uma decisão no seu trabalho, o faça utilizando o conhecimento de toda a organização, ao invés de usar somente o de um único indivíduo, mesmo que este seja muito capaz. As configurações de transferência de conhecimento, um-para-um, e um-para-muitos representam apenas a metade do universo de possibilidades. Entretanto, as grandes oportunidades se encontram no campo onde as pessoas e as organizações se sentem menos confortáveis – transferência de conhecimento de muitos para muitos. Podemos considerar que o objetivo da gestão do conhecimento é assegurar que cada pessoa, ao tomar uma decisão, tenha a sua disposição toda a experiência e o intelecto da sua organização. Desta forma é fácil de perceber que os grandes ganhos com a gestão do conhecimento estejam associados à forma de compartilhamento *muitos-para-muitos*. Mas esta é também a forma de compartilhamento mais difícil de ser colocada em prática. Não é a forma intuitiva de se pensar em troca de conhecimento. Há, portanto, barreiras culturais que devem ser suplantadas. Mas aqui reside uma grande oportunidade de alavancar o desenvolvimento estratégico da empresa através do conhecimento.

4.4 Gestão do Conhecimento sob o Prisma da Informação e Tecnologia [30]

Há um inexorável vínculo entre a Gestão do Conhecimento e a questão da informação. Quando a informação é adequadamente capturada, tratada e disponibilizada, dinamiza-se o seu fluxo fazendo com que tenhamos um impacto virtuoso sobre o processo de gestão do conhecimento. Quando os fluxos de informação

[30] Questões relacionadas à Tecnologia da Informação serão abordadas com maior detalhe no Capítulo 7.

são dinâmicos e as tecnologias que os suportam estão adequadas a tal propósito, temos melhor integração entre grupos de pessoas, análises e avaliações mais ágeis dos objetos de conhecimento que estão sendo interpretados, bem como melhor tomada de decisão e atuação de pessoas e organizações. É um desafio hoje dinamizar os fluxos de informação considerando os volumes crescentes delas nos últimos anos. Diante do turbilhão de informações que possuímos, torna-se necessária uma apreciação adequada da mesma. De acordo com BUSH (1945):

> *"as publicações se estenderam muito além de nossa capacidade atual de fazer real uso delas. O registro das idéias humanas expandiu-se prodigiosamente -- e, no entanto, os meios de que nos valemos para tentar encontrar algo de importante nesse labirinto de idéias, são os mesmos que utilizávamos quando muito menos existia para pesquisar".*

A perspectiva do estudo da gestão do conhecimento sob o prisma do suporte informacional e tecnológico está vinculada à análise das teorias, estratégias, conceitos, processos, instrumentos, tecnologias e ferramentas que podem viabilizar o processo de geração, retenção tratamento e utilização de dados e informações como suporte à gestão do conhecimento pela organização e que por isso mesmo envolvem metodologias e tecnologias da informação. A Ciência da Informação tem muito a dizer, então. Conforme BORKO (1968):

> *A ciência da informação é a disciplina que investiga a propriedade e o comportamento da informação, as forças governantes do fluxo de informações e os meios de processamento da informação para otimizar a acessibilidade e usabilidade. Ela engloba um corpo de conhecimentos relativos à origem, coleção, organização, armazenamento e recuperação, interpretação, transmissão, transformação e utilização da informação.*

Por outro lado, a Tecnologia da Informação, apoiada na Ciência da Computação e na Engenharia de Computação, tornou-se uma dimensão indispensável para viabilizar a proposta da Ciência da Informação, ao definir e processar algoritmos bem como construir sucessivos e avançados equipamentos (hardware) que assegure resultados computacionais cada vez melhores. Abre-se então uma perspectiva para que uma força interdisciplinar oriunda da conjunção entre a ciência da informação e as tecnologias da informação, especialmente suportadas pelos processos computacionais, possa proporcionar melhores resultados à gestão do conhecimento. Devemos assumir como

necessária e importante a relação entre Ciência da Computação e Ciência da Informação para que, de tal sinergia possa emergir resultados satisfatórios para o usuário. De acordo com SARACEVIC (1992), a base da relação entre Ciência da Informação e Ciência da Computação reside na aplicação dos computadores e da computação na recuperação da informação, bem como nos produtos, serviços e redes associados. Enquanto a ciência da computação trata de algoritmos que transformam informações, a Ciência da Informação trata da natureza mesma da informação e sua comunicação para uso pelas pessoas. Ambos os objetos são interligados e complementares, embora possam ter agendas básicas e aplicadas diferentes. Ainda na perspectiva desse autor, os trabalhos com sistemas inteligentes, bases de conhecimento, hipertextos e sistemas relacionados, interfaces inteligentes e interação homem-computador devem ter uma relação importante com a Ciência da Informação, conforme afirma:

> *Essas áreas tem um significativo componente informacional, associado com a representação da informação, sua organização intelectual e encadeamentos; busca e recuperação de informação; a qualidade,o valor e o uso da informação todos tradicionalmente tratados pela CI. De modo inverso, essa corrente de pesquisa e desenvolvimento na ciência da computação fornece diferentes visões, modelos e enfoques e um paradigma diverso para a pesquisa e desenvolvimento em CI. Portanto, as relações interdisciplinares estão evoluindo em direção a um nível diferente de cooperação intelectual.*

É importante salientar que tecnologia da informação não é sinônimo de gestão do conhecimento. A informática chegou a apresentar ao mercado softwares aplicativos com a pretensão de automatizar a gestão do conhecimento, o que é obviamente de resultado questionável devido à complexidade do processo e à falta de percepção de que há uma separação entre conhecimento e informação. Além disso, sendo a gestão do conhecimento um processo sinérgico, que envolve muitas dimensões, pode-se considerar a denominação "software de gestão do conhecimento" um tanto quanto inadequada ou forçada. Entretanto, não se pode negar a importância do suporte que a tecnologia da informação confere à gestão do conhecimento quando articula múltiplas alternativas de hardware e software para apoiar o processo, suporte esse cada vez mais importante, mas nunca completo e exaustivo.

Considerando a perspectiva de CHOO (2003) constatamos que a informação é um componente intrínseco de tudo o que uma organização faz. Para esse autor, há três

pontos fundamentais onde a criação e o uso da informação desempenham um papel estratégico no crescimento e capacidade de adaptação da empresa. Primeiro, a organização usa a informação para criar significado, dando sentido às mudanças do ambiente externo. Segundo, a organização cria, organiza e processa a informação de modo a gerar novos conhecimentos por meio do aprendizado. Novos conhecimentos permitem a criação de novos produtos e serviços ou aperfeiçoar os já existentes, bem como os processos organizacionais. Terceiro, as organizações buscam e avaliam informações de modo a tomar decisões importantes. Esses três pontos são processos interligados e assim devem ser vistos, garantindo uma perspectiva holística do processo.

Para CHOO (2003) a informação é, em essência, a matéria prima do cotidiano do ser humano. A busca da informação é um processo pelo qual o individuo procura obter informações com um propósito definido, buscando modificar o seu estado de conhecimento. A informação é utilizada para que a pessoa possa dar sentido a uma experiência, bem como de agir ou reagir à luz de um novo conhecimento. O comportamento das pessoas que buscam a informação passa pelo crivo da situação (as pessoas constatam a situação em que a necessidade da informação é configurada), da cognição (preenchem lacunas de conhecimento) e da emoção (percebem que o estado psicológico e a emoção afetam a busca da informação). Portanto, a informação precisa ser enfocada sob o ponto de vista do comportamento humano, envolvendo três dimensões: a cognitiva, a emocional e a situacional. Segundo ele,

> *"no nível cognitivo, diferentes estratégias de busca de informações são ativadas para preencher diferentes lacunas de conhecimento. No nível afetivo, o estado emocional e psicológico determina diferentes preferências e métodos de buscar a informação. No nível situacional, as características do trabalho ou da situação problemática determinam a maneira de usar e acessar a informação* (CHOO, 2003)*".*

Do ponto de vista estrutural, a informação precisa ser avaliada no que tange à criação de significados, a criação de conhecimento e o processo de tomada de decisões. Na realidade, CHOO (2003) trabalha a perspectiva de como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. Para esse autor, as chamadas "Organizações do Conhecimento" utilizam de modo estratégico a informação para três perspectivas de ação distintas mais interligadas, quais sejam:

- Criação de significado (*sensemaking*)
- Criação de conhecimento por meio da aprendizagem organizacional
- Tomada de decisão com base no princípio da racionalidade limitada de Herbert Simon[31]

Na etapa de criação de significado, CHOO (2003) postula que o objetivo imediato do "*sensemaking*" é permitir aos membros da organização a construção de um entendimento compartilhado do que é a organização e o que ela faz. Busca-se a sobrevivência e prosperidade da empresa no longo prazo e em um ambiente em constante transformação. Por isso, é necessário interpretar o ambiente, analisando e conferindo sentido aos dados coletados desse ambiente. A inteligência competitiva e o monitoramento ambiental sistemático são práticas organizacionais de construção de sentido. Na visão desse autor:

> *"A criação de significados relaciona-se com a captação de mensagens sobre o ambiente. É necessário identificar o que acontece, dar sentido aos acontecimentos e, posteriormente, desenvolver uma interpretação comum para direcionar uma ação, uma decisão.*

[31] De acordo com esse autor, toda a racionalidade no processo de decisão é limitada. O gestor não maximiza, toma decisões que o satisfazem, descobre soluções aceitáveis para problemas bem reais. Contenta-se com alternativas satisfatórias. Herbert Simon, Prêmio Nobel da Economia de 1978, criticou a abordagem neo-clássica da gestão que encara o processo de decisão por parte do gestor ou do empresário como algo "cartesiano", racional, visando a otimização e maximização de objetivos bem delineados. Uma das contribuições de Herbert Simon para a Teoria da Administração foi a contestação desta visão. De acordo com esse autor, para se compreender a gestão, tem de se perceber como é que as pessoas realmente resolvem problemas e tomam decisões. Há enormes limitações na capacidade de uma pessoa levar em consideração, para a sua decisão, todos os fatos que se dão no mundo, ou em seu ambiente, e que seriam relevantes para essa decisão. Estes limites - a racionalidade limitada - surgem simplesmente porque os seres humanos têm um conhecimento restrito, têm capacidades limitadas para poder analisar as conseqüências do próprio saber que elas detém. Especialmente, há sérios limites para poder predizer o futuro e as reações dos outros às suas decisões. Portanto, não é possível saber todas as alternativas que estão disponíveis para a tomada de decisão - muitas destas próprias oportunidades estão por se descobrir ou desenhar, e as decisões têm de ser tomadas, apesar de tudo, muitas vezes considerando uma ínfima fração daquelas possibilidades. Assim sendo, quando alguns economistas afirmam que «as pessoas racionais "atuam no sentido de maximizar", não só falam de algo que não se "encaixa" com fatos observáveis, como não ajudam em nada a que se tomem boas decisões – ou seja: decisões que possam resolver problemas do mundo real que enfrentamos, mesmo que estas soluções nunca sejam "ótimas" no sentido literal do termo. As decisões que tomamos "satisfazem", não otimizam. Herbert Simon – Entrevista. Herbert Alexander Simon tem uma produção teórica invejável em vários domínios. Na área da economia e da gestão há duas obras suas consideradas fundamentais: "*Administrative Behavior*", publicada em 1947, e "*Models of Bounded Rationality*" (sobre o tema da racionalidade limitada, que lhe deu celebridade), cujos dois primeiros volumes foram escritos em 1982, já depois de laureado com o Nobel da Economia. Em http://www.janelanaweb.com/digitais/simondigest.html. Acessado em 10 set 2006

Segundo CHOO (2003), a etapa de criação do conhecimento é o momento em que as organizações criam ou adquirem, organizam e processam a informação com o propósito de gerar novo conhecimento através da aprendizagem organizacional. O novo conhecimento gerado permite que a organização desenvolva novas habilidades e capacidades, crie novos produtos e serviços, aperfeiçoe os antigos e melhore seus processos organizacionais. Em suas palavras:

> *A construção do conhecimento é, em essência, o adequado relacionamento do conhecimento tácito e do conhecimento explícito pessoal ou organizacional, visando a criação de mais conhecimentos.*

Já no que tange à tomada de decisões, considerando a racionalidade limitada já citada, CHOO (2003) afirma que:

> *A tomada de decisões é uma etapa natural depois da criação de significados e da construção do conhecimento. Nesse momento, definem-se as alternativas possíveis para tomar a decisão que, em tese, deveria envolver a três dimensões do psiquismo humano, contribuindo para miminizar conflitos, tanto para o tomador de decisões quanto para o grupo sujeito à tomada de decisão".*

Cumpre salientar que além do conhecimento tácito e do explicito, CHOO caracteriza também o conhecimento cultural, expresso nas crenças, valores, normas e pressupostos utilizados para dar valor a novos conhecimentos e informações.

O prisma do suporte informacional e tecnológico será desenvolvido nos próximos capítulos, abordando os vínculos entre recuperação da informação, sistemas de descoberta de conhecimento, tecnologia da informação e gestão do conhecimento.

4.5. Aproximações Conceituais dos Prismas Cognitivo e Organizacional

A partir do exposto até o momento, podemos inferir algumas observações sobre o conhecimento na perspectiva teórica de PIAGET (1976). Nela, o conhecimento é expresso pelas estruturas de representação da realidade, que vão sendo construídas pelo indivíduo ao longo do tempo através da interação deste com o seu meio. Essas estruturas se modificam e se auto-regulam no que foi descrito como processo de

equilibração majorante. Esse processo é compatível com teoria do conhecimento tácito e explícito de POLANYI (1969).

Se assumirmos a perspectiva de PIAGET, na tentativa de aplicá-la ao universo da organização, podemos sugerir que também ela possui suas estruturas de interpretação e representação da realidade, ou seja, conhecimento. Esse conhecimento se expressa pela cultura, pelos valores e princípios, pela ação estratégica e pela "memória organizacional" (presente nos bancos de dados e sistemas de informação). Também o conhecimento organizacional evolui na medida em que a organização interage com o ambiente, assimila elementos da realidade e se acomoda a ela. Esse processo emerge das relações entre os indivíduos no ambiente de trabalho quando compartilham suas próprias estruturas individuais de representação (ou conhecimento) entre si, no seio da cultura a qual, ao mesmo tempo em que os influencia, é influenciada por esses mesmos indivíduos (nesse contexto, empregados de uma organização).

Analisando as perspectivas propostas por PIAGET (1976) e POLANYI (1969) e confrontando com o prisma organizacional que tem foco em NONAKA & TAKEUCHI (1997) e DAVENPORT & PRUSAK (1998) podemos perceber aproximações explícitas e subjacentes. Dentre as aproximações explícitas podemos citar os conceitos de conhecimento tácito e explicito de POLANYI (1969) que são utilizados pelos dois grupos de autores organizacionais. Por outro lado, podemos avaliar as aproximações subjacentes, como entre NONAKA & TAKEUCHI (1997) e PIAGET (1976): confrontando a teoria de criação do conhecimento organizacional do primeiro com a teoria da equilibração majorante perceberemos que, embora se refira à organização ou a pessoas no contexto organizacional, ambas parecem ter o mesmo arcabouço. Os processos de socialização, externalização, combinação, interiorização da criação do conhecimento organizacional se confrontam bem com os de ação, assimilação, organização, acomodação e estruturas de representação de PIAGET (1976). Ou seja, as idéias desse último parecem estar subjacentes ao núcleo teórico de NONAKA & TAKEUCHI, os quais, em suas colocações parecem extrapolar o desenvolvimento teórico de PIAGET, focado na construção do conhecimento pelo indivíduo para um processo análogo na organização.

CANNON-BROWERS, J.A ., E. SALAS, e S. CONVERSE (1993), citados por NONAKA & TAKEUCHI (1997), ao descreverem o conceito de modelos mentais compartilhados, tem como linha subjacente a teoria de PIAGET (1976), inovando, no entanto ao aplicá-la às pessoas trabalhando em conjunto.

Também podemos fazer uma aproximação entre as afirmações de DAVENPORT & PRUSAK (1998: 90) sobre mapas do conhecimento individuais e organizacionais com as mesmas idéias das estruturas de representação do conhecimento de PIAGET (1976). Percebemos a convergência das idéias e uma mesma raiz que indica o conhecimento como um estado do desenvolvimento individual e que pode ser compartilhado com outras pessoas para termos então um "conhecimento organizacional"

Poderemos identificar outras aproximações da visão de PIAGET com outros autores do campo organizacional que realizaram enfoques sobre questões de aprendizagem nas organizações (embora para PIAGET aprendizagem e conhecimento sejam elementos diferentes), mas que preferimos não tratar nessa dissertação em função do seu escopo e de suas limitações[32].

4.6. O Prisma do Fator Humano no Conhecimento

A Gestão do Conhecimento é um campo interdisciplinar que além dos aspectos da informação e tecnológicos já apresentados, deve também se articular com o campo da Administração. A Administração é o campo do conhecimento que tem as organizações por objeto de estudo e como tal também estuda o comportamento humano dentro dessas mesmas organizações como fator relevante para elas. É importante salientar que as pessoas possuem suas visões de mundo, seus valores e seus interesses e mesmo dependendo da venda de seu trabalho para sobreviver, de alguma forma procuram afirmar isso de múltiplas formas. Possuem, também, suas habilidades, seus

[32] Como exemplo podemos citar: ARGYRIS, C. *Doublé Loop Learning Organizations*. Harvard Business Review (set-out 1977: 115-125). Também SENGE, P. A Quinta Disciplina – arte, teoria e prática da organização de aprendizagem. São Paulo: Best Seler, 1990. De igual forma: KIM, D. *The Link between Individual and Organizational Learning*. Sloan Managemente Review, (out. 1993: 37-50).

talentos, sua *expertise*. Tais aspectos podem estar ou não vinculados a uma diretriz que valorize e estimule práticas de gestão do conhecimento.

Por outro lado, de nada adianta todo o aperfeiçoamento tecnológico e a melhoria dos fluxos de informação se as pessoas não são capazes de fazer bom proveito de tais avanços. Por mais que avancem as tecnologias e mais sofisticados que sejam hardware e software, dificilmente poderemos prescindir do ser humano como elemento central e decisivo em todo o processo de gestão do conhecimento. Mesmo com o avanço dos sistemas denominados "inteligentes" dentre os quais estão incluídos as propostas de mineração de textos, o ser humano continua sendo elemento decisivo para entender, avaliar e interpretar conjuntos de informações a ele disponibilizados e dar-lhes efetivo sentido. É ele quem pode participar proativamente das *intranets* organizacionais e dá-lhes vigor, fazendo com que efetivamente exista compartilhamento de informação e conhecimento, da mesma forma que é o ser humano quem interpreta indicadores financeiros, analisa e coteja resumos de textos para tomar decisões adequadas para uma organização. Cada vez mais as organizações desejam identificar e capturar o conhecimento estratégico, dentro e fora delas, retendo e protegendo o conhecimento de valor, orquestrando sua consolidação e disseminação para gerar inovações e garantindo o desenvolvimento de vantagens competitivas. A principal vantagem competitiva de uma organização, portanto, é o talento humano e a competência das pessoas que trabalham para ela.

A gestão do conhecimento, sob o prisma do fator humano, articula-se com a gestão por competências, prática da gestão de talentos humanos, cada vez mais presente nas organizações. O que vem sendo chamado de "modelo de competências" em geral implica no desenvolvimento de uma taxonomia vinculada ao processo de gestão de pessoas no contexto do conhecimento, ou então de tentar valorizar o conhecimento que está nas pessoas. A definição de competências envolve:

- Uma combinação de conhecimentos, saber fazer, experiências e comportamentos;
- Exercício num contexto preciso;
- Implica em tomar iniciativa e assumir responsabilidade diante de situações com as quais uma pessoa pode se deparar, incluindo a noção de imprevisto;

- Entendimento prático de situações que se apóia em conhecimentos adquiridos e os transforma na medida em que aumenta a diversidade das situações.

O conceito de Competências, no contexto do campo da Gestão de Pessoas ou das áreas de Recursos Humanos das empresas vem sendo entendido como o conjunto de conhecimentos, habilidades e atitudes. Tais competências configuram o aporte de conhecimento e esforço para viabilizar os processos que a organização possui. Portanto, indicam o potencial humano para gerar as respostas desejadas pela organização. O mapeamento das competências que a organização possui, bem como daquelas que necessita possuir é um processo desenvolvido em projetos de Gestão do Conhecimento. Identificar quais pessoas realizam "o quê", com qual "formação e experiência" e onde elas se encontram física e organizacionalmente é o que se coloca. Estimar a lacuna entre competências atuais possuídas e as almejadas, gerando uma classificação e um índice para localização é um esforço legítimo de organizações preocupadas com o crescimento, a inovação e a gestão de processos complexos de operação e negócio.

O Fator Humano é um prisma de estudo sobre gestão do conhecimento que tem um grande campo para desdobramentos e a presente tese apenas o pontua. No contexto da pesquisa desenvolvida, as pessoas têm um papel decisivo para dinamizar os resultados das implantações e aplicações de novas tecnologias, como é a proposta de estudo que propõe o uso da mineração de textos, cujo processo, uso e interpretações e avaliações de seus resultados serão feitas por pessoas.

Um conjunto de visões, metodologias e procedimentos, vinculados às práticas gerenciais da chamada Administração de Recursos Humanos – ARH ou, mais modernamente chamada de "Gestão de Pessoas" estão sendo usadas por organizações para apoiar e viabilizar iniciativas e programas de Gestão do Conhecimento, tais como: gestão por competências, mapeamento de competências, gestão de carreiras, *mentoring*, implantação de universidades corporativas, uso de mapas conceituais, dentre outras. Programas de treinamento estruturado para cobrir "*gaps*" (espaços) de conhecimento mapeados e para fornecer conhecimento complexo são também utilizados como ações de ARH para subsidiar a Gestão do Conhecimento, cujo papel em uma organização pode ser sintetizado nos itens seguintes:

- estabelecer juntamente com as áreas técnicas e operacionais os perfis necessários a execução das diversas atividades;
- coordenar um processo de mudança cultural para criar um ambiente adequado a criação e o compartilhamento de conhecimento;
- providenciar programa de treinamento para reduzir as lacunas identificadas no mapeamento dos capitais do conhecimento;
- organizar plano de substituição do pessoal (plano de sucessão), e programas de treinamento no trabalho (mentoring);
- desenvolver e implementar uma política de adequada de Recursos Humanos.

4.7. Uma apreciação crítica à gestão do conhecimento

A idéia de Gestão do Conhecimento não é um consenso. Há muitas críticas e questionamentos relevantes são a ela formulados. NEHMY (2001) analisa que o programa de gestão do conhecimento pretende abordar o conhecimento numa dimensão micro-social, aquela relativa às organizações e define-o da seguinte forma:

> *"Programa de tecnologia gerencial que propõe forma alternativa de otimizar os recursos e a competitividade das empresas no cenário das novas tecnologias e da globalização. É sedutor. Uma característica importante desse programa e que o diferencia de outros modismos da Administração é sua aproximação do formato acadêmico, valorizando o arcabouço teórico e desenvolvendo-se também como programa de pesquisa, o que aumenta suas condições de institucionalização e disseminação no mundo acadêmico. Mesmo assim, esse programa não pode ser caracterizado como científico, uma vez que se encontra em um momento inaugural e devido ao fato das tecnologias gerenciais serem por demais efêmeros. O melhor é caracteriza-lo como um movimento, uma tendência emergente do mundo dos negócios".*

A própria definição de conhecimento na sociedade contemporânea vem sendo fortemente influenciada pelos ventos do mercado. A esse respeito, KURZ (2002) afirmará:

> *"Conhecimento é poder — trata-se de um velho lema da filosofia burguesa moderna, que foi utilizado pelo movimento dos operários europeus do século 19. Antigamente conhecimento era visto como algo sagrado. Desde sempre homens se esforçaram para acumular e transmitir conhecimentos. Toda sociedade é definida, afinal de contas, pelo tipo de conhecimento de que dispõe. Isso vale tanto para o conhecimento natural quanto para o religioso ou para a reflexão teórico-social. Na modernidade o conhecimento é representado, por um lado, pelo saber oficial, marcado pelas ciências naturais, e, por outro, pela "inteligência livre-flutuante" (Karl Mannheim) da crítica social teórica. Desde*

o século 18 predominam essas formas de conhecimento. Mais espantoso deve parecer que há alguns anos esteja se disseminando o discurso da "sociedade do conhecimento" que chega com o século 21; como se só agora tivessem descoberto o verdadeiro conhecimento e como se a sociedade até hoje não tivesse sido uma "sociedade do conhecimento". Pelo menos os paladinos da nova palavra-chave sugerem algo como um progresso intelectual, um novo significado, uma avaliação mais elevada e uma generalização do conhecimento na sociedade. Sobretudo se alega que a suposta aplicação econômica do conhecimento esteja assumindo uma forma completamente diferente".

A principal vertente crítica aponta na direção que o movimento de gestão do conhecimento situa-se na perspectiva de desdobramento contemporâneo da dinâmica capitalista, que cria a cada tempo histórico, novas manifestações e formas de se apresentar. IANNI (1992, p. 47) irá afirmar**:**

"O capital dissolve, recobre ou recria formas de vida e trabalho, de ser e pensar, em âmbito local, regional, nacional e internacional. Simboliza uma espécie de revolução burguesa permanente, ainda que desigual e contraditória, progressiva e regressiva, democrática e autoritária".

O desejo de obter resultados expressivos sobre o investimento realizado e as pressões da concorrência configura o capitalismo como um processo dinâmico no que diz respeito ao desenvolvimento das forças produtivas. A competição econômica força as empresas a buscarem a eficácia, modificando o processo de trabalho em sua estrutura, criando novos produtos e impondo um novo modo de se trabalhar. O processo de desenvolvimento do capitalismo aprofunda a relação já existente entre desenvolvimento científico e tecnológico e aprimoramento das forças produtivas. Na perspectiva de KURZ (1993), há um aprofundamento desde processo no decorrer do século XX, sobretudo em suas últimas duas décadas, com um casamento cada vez mais sólido entre desenvolvimento de novas tecnologias e produção capitalista. Isso é explicado pela lógica do sistema mundial de produção de mercadorias buscar sempre maior eficiência e produtividade. Portanto, é difícil abordar o funcionamento das economias capitalistas sem considerar o desenvolvimento tecnológico. E desenvolvimento tecnológico está intrinsecamente ligado ao processo pelo qual o conhecimento é gerado e alocado em desenvolvimento das forças produtivas. Captar o entendimento de como a tecnologia afeta a economia é decisivo para a compreensão do crescimento da riqueza dos países e dinâmica das sociedades contemporâneas. SMITH (1983) reforça a importância da mudança técnica. MARX (1968) aponta o capitalismo

como um sistema em que o progresso tecnológico é intrinsecamente produzido. SCHUMPETER (1985) fixa a inovação tecnológica no centro da dinâmica do capitalismo. Inovações modificam a realidade econômica e social, além de aumentarem a capacidade de acumulação de riqueza e geração de renda. Portanto, qual a relação entre conhecimento e a produção contemporânea?

Podemos dizer que toda a discussão recente em torno do papel do conhecimento liga-se à questão da ampliação da competitividade, num ambiente no qual a inovação e o desenvolvimento de novos produtos se tornou fundamental para a sobrevivência dentro do sistema, uma vez que: é necessário inovar para garantir maiores vendas a uma base de consumidores cada vez menor e mais exigente. O ciclo de vida dos produtos diminui e eles precisar ser redesenhados a cada instante para se tornar mais atrativos ao consumidor. MÈSZÁROS (2002) discute a questão da obsolescência planejada e da "caduquice" dos produtos como sendo algo necessário às empresas modernas, uma vez que, num ambiente onde a exclusão ao consumo é crescente, somente assim é possível manter níveis de vendas que interessam e garantam a sobrevivência das empresas redesenhadas pela tecnologia que, ao excluir tantos trabalhadores, excluiu também muitos consumidores, dificultando o funcionamento do mercado, o que caracteriza uma grande contradição histórica do capitalismo.

O conhecimento sempre foi importante para os sistemas produtivos. Seu papel está fortemente ligado à produção de inovações. O que podemos perceber hoje é que, em função de uma maior velocidade dos processos de mudanças, ele se tornou ainda mais importante. Não que exista uma mudança substantiva em curso no sistema social ou econômico. A dinâmica capitalista persiste e aprofunda seus mecanismos de geração de excedentes e persistindo os processos de propriedade privada, geração de mais valia e a existência de duas categorias bem distintas que são os capitalistas e os trabalhadores assalariados. O que muda é o fato do sistema produtivo ter absorvido e continuar absorvendo doses elevadas de tecnologias que modificam severamente os processos de produção. Um bom exemplo disso é a intensificação da microeletrônica nos sistemas produtivos. O modelo taylorista-fordista e a abordagem de automação de base eletromecânica não conseguiram fazer o que a microeletrônica vem conseguindo no aprimoramento dos processos de automação A introdução crescente na indústria da automação de base microeletrônica permitiu que as atividades de produção discreta

pudessem também ser alvo de aumentos de produtividade e redução dos custos de mão-de-obra.[33]

Diante do exposto, uma crítica importante ao programa de gestão do conhecimento é sua intenção de apresentar este objeto como sendo uma novidade nas estratégias de competitividade empresarial, o que não é verdade. O discurso do programa de gestão do conhecimento pretende transformar o conhecimento dos trabalhadores em principal recurso econômico, dando muitas vezes a isso o caráter de novidade. Entretanto, o conhecimento dos trabalhadores sempre foi o fator de produção mais importante, seja através do trabalho vivo, seja através do trabalho morto, com cristalização desse conhecimento sob a forma de processos automatizados e outras tecnologias de produção (NEHMY, 2001). Sob esse aspecto, é importante salientar que todo esforço *taylorista* caminhou nessa direção. A ação de Taylor no início do século XX se expressou pela busca de quebrar a "redoma de vidro" que possuía o trabalhador e, ao desvendar as nuances do trabalho, extrair o conhecimento que o operariado detinha sobre o processo de trabalho. A idéia de redoma de vidro é proposta por MORAES NETO (1987) ao estudar a dinâmica entre conhecimento e trabalho na ação promovida por Taylor no inicio do século XX. Segundo MORAES NETO (1987) Taylor vai quebrar a redoma de vidro que envolve o trabalhador da manufatura no seu tempo. Este trabalhador possuía uma forte expertise, um "saber fazer" relevante, mas que era pouco compreendido pelo empresário capitalista. A expertise dos trabalhadores sofreu a tentativa (realizada com êxito) de explicitação por parte de Taylor, com sua representação através de normas, regulamentos e definição dos processos e métodos. Através dessa captura, Taylor pode interferir fortemente na dinâmica de produção do seu tempo. Por sua vez, a Administração, através das atividades chamadas e Organização e Métodos, intensificada desde 1940, procura conhecer, interpretar e sistematizar o conhecimento da organização, via mapeamento e representação de processos, captando e 'instrumentalizando" o conhecimento embutido nas atividades dos empregados, direcionando esse conhecimento da melhor forma requerida pelos imperativos da produtividade, racionalidade e redução de custos. FOUCAULT (1975) citado por (NEHMY, 2001), vai estabelecer uma íntima relação entre poder e sistemas

[33] A automação eletromecânica foi eficiente em grandes campos para os processos de produção contínua. A microeletrônica, além de romper novas barreiras para a própria produção de fluxo continuo, vai permitir a ruptura de barreiras orgânicas na produção discreta, fazendo com que atividades antes impossíveis de serem automatizadas possam, nessa nova abordagem, vir a ser.

de conhecimento, na medida em que embutem técnicas e práticas para o exercício do controle e domínio social em contextos localizados e particulares (a família, a escola, a prisão, o hospital, a empresa).

Outra crítica importante diz respeito ao papel do novo trabalhador na proposta "sociedade do conhecimento". Para KURZ (1993) temos a eclosão de um capitalismo sem trabalho, uma vez que a microeletrônica permitiu romper barreiras orgânicas de produção que antes não eram imaginadas. O aprofundamento da relação entre a ciência e o capitalismo, embora esteja dentro da lógica de busca do desenvolvimento das forças produtivas, provoca agora um colapso histórico desencadeando a crise da sociedade do trabalho. A quantidade de trabalhadores requeridos ao processo de trabalho passa a ser cada vez menor, uma vez que a automação de base microeletrônica amplia as fronteiras das possibilidades de uso nos setores industriais e de serviços. Portanto, um aspecto importante que merece destaque aqui é, por mais contraditório que possa parecer, é o problema da desqualificação. Em muitos setores industriais temos, além da redução quantitativa do número de trabalhadores, a diminuição real da necessidade de qualificação. A operação de processos produtivos com intensidade de automação, em geral, requer com certeza grandes níveis de qualificação no seu planejamento e gestão, mas cada vez menor qualificação na operação, minimizando, portanto a importância do conhecimento. Qual será a importância do conhecimento[34] para uma operador de máquinas com CNC – comando numérico computadorizado, um operador de centrais de atendimento (*call center*) ou um trabalhador bancário nos dias atuais? A mediação da tecnologia, ao contrário de boa parte do discurso, é desqualificante e a importância do conhecimento é, para a maior parte dos trabalhadores, uma quimera. Muitos trabalhadores se transformam em apêndices de máquinas, equipamentos ou sistemas informatizados e cuja operação requer um conjunto de conhecimentos inferior ao que se tinha no passado. Por conclusão temos que, o processo de intensificação tecnológica dos nossos dias sofistica e aprofunda a lógica taylorista-fordista de expropriação do conhecimento do trabalhador e da banalização da presença do mesmo no processo produtivo. Obviamente iremos perceber trabalhadores envolvidos em processos de gestão ou de conhecimento intensivo, como produção de softwares ou desenvolvimento de novos produtos, mas numa proporção muito menor do que na categoria explicitada.

---

[34] E qual tipo de conhecimento? Em geral requer-se o conhecimento operacional ou instrumental, voltado às necessidades diretas de produção.

Uma terceira crítica importante diz respeito ao exagero com que é reforçada a questão da tecnologia informação em muitas experiências, como se elas pudessem resolver todos os problemas de captura, organização, disponibilização e uso da informação. Essa informação, de certa maneira é vista como sinônimo de conhecimento. Muitas empresas desenvolveram softwares em quantidade crescente para outras que adquiriram os mesmos na esperança de ampliarem as fronteiras de gestão e uso do conhecimento (ou seria da informação, uma vez que não há ainda consenso sobre se temos na verdade muito mais informação do que conhecimento efetivo). NEHMY (2001) aponta que o programa de gestão do conhecimento pretende abordar o conhecimento numa dimensão micro-social, aquela relativa às organizações. Para essa autora, o fenômeno da "gestão do conhecimento" possui forte imprecisão teórica, sendo as formulações dos seus autores principais ainda muito frágeis. KURZ (2002) critica severamente esta idéia de explosão de informação ao afirmar:

> *"Elucidativo é talvez o fato de que o conceito da "sociedade do conhecimento" esteja sendo usado mais ou menos como sinônimo do de "sociedade da informação". Vivemos numa sociedade do conhecimento porque somos soterrados por informações. Nunca antes houve tanta informação sendo transmitida por tantos meios ao mesmo tempo. Mas esse dilúvio de informações é de fato idêntico a conhecimento? Estamos informados sobre o caráter da informação? Conhecemos afinal que tipo de conhecimento é esse? Na verdade o conceito de informação não é, de modo nenhum, abarcado por uma compreensão bem elaborada do conhecimento. O significado de "informação" é tomado num sentido muito mais amplo e refere-se também a procedimentos mecânicos. O som de uma buzina, a mensagem automática da próxima estação do metrô, a campainha de um despertador, o panorama do noticiário na TV, o alto-falante do supermercado, as oscilações da Bolsa, a previsão do tempo... tudo isso são informações, e poderíamos continuar a lista infinitamente".*

Um quarto aspecto de crítica à questão da gestão do conhecimento é a idéia de empresa do conhecimento que valoriza o papel da inovação e, portanto, da geração do conhecimento necessário para produzi-la. Não há como comparar trabalhadores do conhecimento com cientistas em uma comunidade acadêmica. No ambiente acadêmico, a autoria, a posse e o reconhecimento garantem ao cientista uma relação digna e direta entre o seu trabalho e os frutos do conhecimento produzido, enquanto na empresa temos que a intenção desta é se apropriar e utilizar desse bem do trabalhador para transformá-lo em bem e estoque da empresa. Para NEHMY (2001):

> *"Quando o trabalhador expõe o seu conhecimento, e o socializa, o conhecimento não se acumula, ao reverso, perde-se. Não é possível provocar adesões fortes a essas propostas. O conhecimento racional, lógico esperado do trabalhador detentor do saber pessoal especialista teria de ser de resistência e não de predisposição favorável(...) Fica bem clara na exposição dos diferentes autores a não existência da transferência 'espontânea'do conhecimento do trabalhador nas empresas, indicando a necessidade alavanca (um motor exterior ao cotidiano da vida da empresa) capaz de iniciar e manter o processo de conversão (NONAKA E TEKEUCHI, 1998) Entretanto, o conhecimento do trabalhador sempre foi apropriado pela empresa desde o sistema de administração taylorista".*

Na esteira dessas considerações podemos dizer que temos a eclosão de um quinto ponto de crítica, também adequadamente formulada por NEHMY (2001), que é o fato do ponto de vista do trabalhador ser completamente desconsiderado nas exposições sobre capital intelectual ou gestão do conhecimento, ao afirmar:

> *"A máxima do regime pós-fordista prega a autonomia do trabalhador e a competência pessoal adquire importância primordial, que na gestão do conhecimento torna-se o foco da gerência. (...) O programa de gestão do conhecimento pretende que o trabalhador entregue seu saber técnico, sua experiência (o saber como, e o saber ser) além daqueles aspectos já apropriados pela administração flexível (horas de trabalho para além do horário, o trabalho em casa, participação em reuniões). Nada se fala do crescimento pessoal do trabalhador dentro da empresa. Em nenhum momento o trabalhador é chamado a refletir sobre seu conhecimento no sentido libertador do qual fala Habermas (1982) de forma a amplia-lo e deslocá-lo da situação particular da ação instrumental para constituir o conhecimento pleno, abstrato".*

Na visão de HABERMAS (1982), a reflexão refere-se ao ato de se posicionar face a algo que de alguma maneira não lhe está disponível. Refletindo, damos conta daquilo que pensamos saber. Para reflexão é estranha a idéia de que alguém não possa saber como e porque sabe. Para quem reflete seria *non sense*, por exemplo, acertar o resultado de uma operação matemática sem saber matemática. Nessa dinâmica, o saber fica descomprometido com essa ou aquela experiência singular, e livre para assimilar novas situações, constelações imprevistas, apenas hipotéticas em relação ao que já foi experimentado como conhecimento. Até que ponto podemos realmente considerar o conhecimento como algo que pode ser separado do seu autor? Até que ponto podemos realmente explicitar o conhecimento?

4.8 Avaliação e Posicionamento Sobre as Críticas à Gestão do Conhecimento

Todas essas críticas são importantes para garantir que tanto a reflexão teórica quanto as práticas organizacionais possam aprofundar sua densidade e consistência, evitando percepções incompletas ou muito ideológicas e buscando garantir um enfoque mais adequado da participação e envolvimento das pessoas no que podemos assumir como gestão do conhecimento. Isso é importante também para identificar qual deve ser o verdadeiro papel da Ciência da Informação e da Tecnologia da Informação em suas relações com o conhecimento humano. Hoje, não podemos considerar mais que a questão da Gestão do Conhecimento se encontra em um momento inaugural, uma vez que muitas organizações públicas e privadas já vivenciaram programas em sua esfera, de diferentes tipos e diferentes objetivos.

Embora, conforme foi apontado nas críticas, seja mesmo impossível comparar trabalhadores do conhecimento com cientistas em uma comunidade acadêmica, uma vez que no meio acadêmico cientistas tem o reconhecimento da autoria e a posse do conhecimento e na empresa os trabalhadores possam mesmo ser expropriados do conhecimento que geraram, não necessariamente precisa ser assim. Com o aprofundamento das necessidades de inovação em produtos e processos, estimular e envolver o trabalhador em processos de geração do conhecimento tornou-se fundamental. Hoje vivemos numa sociedade pluralista em que, não obstante a mídia não seja uma instância democrática (uma vez que suas principais expressões são controladas e mantidas por grandes e concentrados grupos econômicos), o trabalhador em muitos paises, tem possibilidades mais amplas de participação na vida social e econômica da nação, embora tenha se fragilizado enquanto classe social pela reestruturação produtiva que fez o sistema capitalista realmente demandar menos trabalhadores e enfraqueceu suas instancias de representação: os sindicatos. Porém, os trabalhadores que continuam a atuar nos contextos empresariais, tendem, em grande parte, ser mais qualificados e conscientes de seu papel, tendo as organizações mais desafios estratégicos em envolvê-los em atitudes colaborativas e de partilha de seu conhecimento, para o qual precisará ser estimulado, de forma cada vez mais consciente e, muitas vezes, com compensação financeira por sua produção intelectual, seja por prêmios, bônus, salários indiretos, etc. Ou seja: embora possa acontecer, talvez até na maior parte das situações, não obrigatoriamente o trabalhador precisa ser ou será expropriado de seu conhecimento

sem o devido reconhecimento e estímulo, mas tal conhecimento pode ter, inclusive, valor econômico. O reconhecimento do valor desse conhecimento e a inclusão do mesmo entre seus pleitos é um novo desafio para as classes trabalhadoras naquilo que vem sendo chamada (adequadamente ou não) de "Sociedade do Conhecimento", ou um novo tempo no qual tal questão passou a ser discutida objetivamente.

Realmente, a questão do valor do conhecimento não é algo novo em importância no contexto econômico, conforme afirmado por KURZ (2002), mas há hoje uma intensificação de sua absorção no processo produtivo das empresas, em busca de maior velocidade na inovação e de maior diferenciação. Nesse sentido, podemos assumir como real a proposta de IANNI (1992) de que:

> *"O capital dissolve, recobre ou recria formas de vida e trabalho, de ser e pensar, em âmbito local, regional, nacional e internacional".*

O que percebemos agora, entretanto, é que esse capital assume uma postura de conferir um valor muito alto para o conhecimento, o que alguns autores abordam como capital intelectual. Para STEWART (1998), o valor de uma empresa é determinado mais pelo seu capital intelectual do que por seu capital financeiro. Alguns autores calculam que 2/3 dos resultados financeiros de uma empresa são devidos ao Capital Intelectual (TERRA, 2001). Segundo STEWART (1998), Capital Intelectual é a soma de 3 capitais:

a) Capital humano: inclui conhecimento e competências individuais dos funcionários (refere-se ao valor que a empresa perde quando funcionários vão embora);
b) Capital estrutural: inclui conhecimento ou competência coletiva, como processos, *know-how*, marcas e patentes, documentos (é o valor que fica quando funcionários vão embora);
c) Capital do cliente: inclui conhecimento e vantagens advindas dos clientes (é o valor que se ganha com relacionamentos com clientes).

4.9 Redes Sociais e Redes de Conhecimento

Na perspectiva de JARVENPAA & TANRIVERDI (2003) o relacionamento "intra" e "interorganizacional", visando a cooperação, é, hoje, ponto central de uma nova forma organizacional que é a rede, uma vez que na rede informação e conhecimento podem melhor fluir entre os atores envolvidos. As unidades que integram a rede, formando um todo, são responsáveis por compartilhar informação e conhecimento na rede. Esse é um assunto vasto e complexo, com muitos desdobramentos e é aqui apenas citado para pontuar a sua relevância, uma vez que é referência para aplicação de certas iniciativas em Gestão do Conhecimento na ELETRONUCLEAR que estão fora do escopo dessa tese, mas que são tangenciadas pelas abordagens e tecnologias aqui estudadas.

4. 10 Conhecimento e Memória Organizacional

Um desafio organizacional significativo, diretamente ligado às práticas de Gestão do Conhecimento e relevante no contexto da presente tese é a questão da administração da memória organizacional. A expressão "memória organizacional" pode gerar polêmica. Tal como "conhecimento organizacional" não há consenso. Alguns podem não concordar com ela. Não há a pretensão de defini-la aqui de modo exaustivo. Por memória queremos dizer mais do que base de dados ou repositórios de documentos. Queremos nos referir a uma linha relevante de articulação de conhecimentos ao longo de um período de tempo os quais tem significado e relevância para uma organização, sendo que tais conhecimentos têm impacto sobre seu agir no presente e no futuro. Eventos do passado da organização ou de seu ambiente podem ter novas releituras e ajudar a solucionar problemas e tomar decisões no presente e no futuro. Promover o dinamismo da memória e o acesso à história, criando uma adequada taxonomia, utilizando tesauros[35], portais corporativos[36], e os processos de recuperação da informação embutidos na tecnologia de mineração de textos abordada nessa Tese torna-

---

[35] Tesauro pode ser definido como uma estrutura de dados composta de uma lista de palavras importantes dum dado domínio de conhecimento e para palavra da lista, um lista de palavras relacionadas (sinónimos, etc...).

[36] Portal Corporativo: Tipo de software que permite configurar uma rede de computadores de uma organização (*intranet*) de forma dinâmica no sentido de integrar distintos repositórios de conteúdos de informação sob um mesmo ambiente de acesso a eles. Através de um portal, desde que esteja devidamente credenciado, o usuário tem acesso a quantos repositórios de dados e documentos que a sua organização possui.

se cada vez mais importante diante dos volumes crescentes de conteúdos que são abarcados por essa memória. ARAUJO (1994), afirma que os sistemas de recuperação de informações dão acesso às informações nele armazenadas e que essas informações constituem a memória humana registrada, o que, segundo ela, BELKIN E ROBERTSON (1976) categorizam como informação “cognitivo-social”: as estruturas conceituais referentes ao conhecimento coletivo, ou seja, as estruturas de conhecimento partilhadas por membros de um grupo social (manuscritos, livros, periódicos, mapas, filmes, vídeos, quadros, partituras, etc.). Podemos acrescentar aqui todos os tipos de documentos digitais.

As organizações sempre vão depender das pessoas e não podem delas prescindir. Entretanto, em grande parte das situações, as organizações possuem objetivos mais amplos que os objetivos individuais e tendem a sobreviver mais tempo do que elas. As pessoas ingressam e deixam as organizações. Durante sua permanência nelas, e no contexto econômico e cultural em que essas pessoas se inserem, muito conhecimento é gerado e grande parte dele pode ser explicitado, sendo traduzido em documentos, informações e dados a ponto de serem armazenados em sistemas de informação computadorizados para serem posteriormente recuperados. Portanto, em certo sentido, impõe-se reduzir a dependência estratégica das pessoas desenhando processos objetivos e institucionais relativos à preservação e recuperação da memória organizacional. Além disso, a consulta a essa memória não é questão hoje de mero exercício de recordação, mas passo preliminar para alavancar novas oportunidades. Despersonalizar os processos, tornando-os mais independentes portanto e facilitar simulação de processos com base no histórico enriquece e aprimora a atividade gerencial. Compartilhar as “melhores práticas”, criar “banco de soluções”, classificar documentos já existentes e incluir novos em portais corporativos e bibliotecas digitais são atividades vinculadas à idéia de gerir a memória da organização.

4.11 Desafios e Perspectivas em Gestão do Conhecimento

A expressão **conhecimento** pode ser definida pelo conjunto de princípios, regras, modelos mentais e memórias nos quais a ação humana está enraizada. Conforme CHOO (2003), o conhecimento pode ser divido em:

- Explícito – aquele conhecimento que pode ser objetivado em dados, informações, memoriais, obras, documentos e mesmo produtos.
- Implícito – aquele conhecimento que está subjacente à ação humana e que emerge da relação do homem com a realidade mas cuja objetivação ainda não se fez ou será dificilmente realizada.
- Cultural – Aquele que está disperso em suposições crenças e valores para descrever e explicar a realidade e para dar valore e importância a cada nova informação.

Somente os indivíduos conhecem, criam coisas novas e podem dizer que geraram novos conhecimentos. Há um ciclo no processo de conhecer perceptível e indicado por:

- aquisição/criação de conhecimento;
- codificação e registro do conhecimento;
- compartilhamento e transferência de conhecimento.

A expressão **conhecimento organizacional** é uma abstração e indica a importância de implementação de modelos que valorizam a sinergia entre processos de criação e transferência de conhecimentos entre pessoas e grupos que nela se encontram. O conhecimento organizacional é um ***status*** do engajamento em grupo das pessoas no sentido expressar sua formação, habilidades, impressões, cultura e comportamentos. A palavra **competência** pode ser usada para objetivar o conhecimento pessoal, que pode ser traduzido em organizacional num determinado momento do tempo. Da mesma forma que o conhecimento não tem limites, também não tem o desenvolvimento das competências.

Embora somente os indivíduos conheçam, a organização pode ser proativa, facilitando, estimulando, orquestrando a sua geração e absorção em inovações de produtos e processos.

O conhecimento organizacional está fortemente objetivado em produtos e processos. Novos produtos e serviços, bem como processos continuamente

redesenhados indicam preocupação com excelência e geração de valor. A inovação tem hoje importância catedral para as organizações. Indica a introdução, na atividade econômica, de novas técnicas, processos, produtos serviços e soluções. Embora seja um fenômeno antigo, o desenvolvimento de inovações, na segunda metade do século XX vai ganhar uma velocidade exponencial. Hoje, podemos dizer que a inovação está no centro das estratégias organizacionais competitivas e portadoras de êxito. Até recentemente, a inovação era fruto de processos de pesquisa e desenvolvimento. Hoje é resultado da aprendizagem organizacional, realidade abstrata que significa em que intensidade os indivíduos, que compõe equipes organizacionais, são capazes de interagirem com a realidade e, coletivamente, produzirem coisas novas e relevantes. Ou, numa perspectiva centrada na teoria da *Equilibração Majorante* de Piaget, atingirem novos e sucessivos patamares de reflexão sobre si mesmas e sobre a realidade, gerando as respectivas competências para nela atuar e transformar. O processo de inovação depende de estruturas de interface entre grupos, pessoas, realidades. Redes de conexões entre seres humanos e organizações, aprendizagem, enfim valoração do que entendemos por conhecimento.

A Gestão do Conhecimento sustenta os processos de criação, transferência e registro do conhecimento, do qual as inovações dependem. Por isso é cada vez mais importante a formalização e a institucionalização do processo de Gestão do Conhecimento nas organizações. A precariedade nos processos de conexão entre equipes e grupos, gerenciamento da aprendizagem e articulação do conhecimento terá por conseqüência baixa criações de inovações. A gestão do conhecimento implicará em ações no campo da gestão das pessoas, dos processos e da tecnologia.

Diante do exposto até o presente momento o conceito de conhecimento de DAVENPORT & PRUSAK (1998: 2-6), o que chamamos de gestão do conhecimento possui, portanto conjuntos distintos de processos relevantes, que devem ser conduzidos com intencionalidade e que podem ser sumarizados nos seguintes tópicos:

- Reconhecimento das diferenças e articulações entre o conhecimento tácito e explícito, conforme as categorias de POLANYI;
- Reconhecimento da importância das teorias do equilíbrio majorante (PIAGET) e da criação do conhecimento organizacional e de suas relações recíprocas;

- Adoção da postura de que todo conhecimento provém da prática social e a ela retorna e de que a busca do conhecimento organizacional é um empreendimento coletivo;
- Percepção da incapacidade e da falta de necessidade de gestão de todo o conhecimento de uma organização, ou seja: o conhecimento que, via de regra, interessa à organização é aquele associado à inovação em produtos, serviços e seus respectivos processos ou à sua gestão e, portanto, essa mesma organização deve delimitar o escopo daquilo que deseja gerir em termos de conhecimento;
- Percepção de que são as pessoas que geram conhecimento e que um possível desenvolvimento do conhecimento da organização se dá a partir do momento em que a cultura dessa organização absorve e é moldada pelo conjunto das múltiplas contribuições e influências das pessoas sobre ela ao longo do tempo o que força sua percepção como um sistema aberto que interage continuamente com o seu ambiente em fluxo contínuo de transformação;
- Percepção de que as pessoas, compartilhando seus conhecimentos umas com as outras, no seio da organização, fazem com que se possa usar uma imagem figurada da organização que gere conhecimento;
- Reconhecimento de que a interação humana, seja ela virtual ou presencial, é fundamental para a construção do conhecimento e que, para muitas circunstâncias a interação presencial é imprescindível;
- Condução consciente do processo de criação do conhecimento organizacional, cuja teoria de NONAKA & TAKEUCHI é bastante elucidativa;
- Construção de uma taxonomia do conhecimento que permita classificá-lo, organizá-lo e recuperá-lo sempre que necessário para a organização.

Conhecimento, Capital Intelectual, Desenvolvimento de Competências, Tecnologia da Informação e *Business Intelligence* são expressões, dentre outras, que passam a fazer parte não só do vocabulário de negócios, mas também da dinâmica organizacional, transformando-se em questões decisivas para a sobrevivência e crescimento de empresas inteligentes e ágeis no ambiente empresarial emergente do milênio que está a iniciar-se. Urge, pois, moldar as condições facilitadoras que capacitam pessoas e empresas a criar e utilizar continuamente o conhecimento em favor da inovação contínua, da diferenciação e do estabelecimento de padrões de excelência na organização, conduzindo processos empresariais que valorizem a Gestão do

Conhecimento, desenvolvendo ações em gestão de pessoas, tecnologia da informação e planejamento geral do modelo de gestão. Pessoas, agindo de forma colaborativa, em times de trabalho e apoiadas por políticas institucionais, metodologias de tratamento do conhecimento e recursos avançados de tecnologia da informação, podem construir uma postura competitiva para as organizações em que trabalham. É possível envolver e articular os times de trabalho na construção constante do conhecimento e na geração de inovações e solução de problemas, com o uso de tecnologia e ações em gestão de pessoas. O desenvolvimento de estratégias organizacionais relacionadas à gestão do conhecimento pode envolver alguns aspectos descritos a seguir.

a) Implementação de modelos conceituais de negócio vinculados a metodologias de tratamento da informação e do conhecimento, como por exemplo, articulação dos conhecimentos tácito e explícito, implicando:

- definição de modelo conceitual de Projeto e Políticas de Manejo do Conhecimento;
- diagnóstico e análise de bases, fontes e dinâmica das articulações entre conhecimento tácito e explicito;
- mapeamento de capital intelectual e Estratégias de Avaliação de Ativos Intangíveis;
- diagnóstico e criação de condições facilitadoras da criação do conhecimento e da inovação;
- desenvolvimento de Modelos de Gerenciamento Baseados tais como *o balanced scorecard.*

b) Desenvolvimento de estratégias e políticas de gestão de talentos humanos que induzam o desenvolvimento do conhecimento na empresa, o que implica em:

- desenvolvimento de modelos e instrumentos de gestão de pessoas por competências;
- estabelecimento de políticas e desenvolvimento de modelos e instrumentos relacionados à educação corporativa e à aprendizagem virtual.

c) Desenvolvimento e implementação de tecnologias da informação e processamento de dados que suportem a demanda por "ativos de conhecimento", tais como *data warehousing, business intelligence* e trabalho colaborativo por computador com *intranets* e a *internet*. Exemplos de ações:

- desenvolvimento de Sistemas de Informações Gerenciais;
- desenvolvimento de Portais Corporativos;
- desenvolvimento de Sistemas Baseados em Inteligência Artificial envolvendo estratégias de mineração de dados;
- desenvolvimento de Sistemas Inteligentes de Pesquisa Textual (Text Mining).

4.12 Conclusão do Capítulo

O papel do conhecimento na vida social é crescente na medida em que se aprofundam as relações de interdependência entre pessoas, organizações e países e na medida em que o êxito das ações humanas e organizacionais depende cada vez mais do bom entendimento da realidade e da capacidade de gerar respostas competentes sobre essa mesma realidade. Seres humanos buscam o sucesso pessoal, organizações aspiram o êxito em seus objetivos e nações anseiam pelo desenvolvimento e crescimento de seus povos e de suas riquezas. Em todas essas dimensões o conhecimento passa a desempenhar papel importante na medida em que está vinculado a palavras chave como o êxito e o sucesso. Esse conhecimento estará em geral ligado à inovação e sua dinâmica implica em constante metamorfose, transformação contínua a partir da construção do saber humano. O conhecimento tem natureza dinâmica e está profundamente caracterizado por dois grandes aspectos: primeiro o mundo humano, das relações entre pessoas que, trabalhando juntas, criam continuamente o conhecimento, solucionando problemas e produzindo inovações. É o conhecimento tácito, tão fundamental quanto difícil de delimitar. Em segundo lugar, é preciso dizer que o conhecimento está ligado a uma importante matéria prima: a informação. Não são a mesma coisa, mas sem a informação, tempestiva, precisa e formatada na necessidade das pessoas, torna-se muito difícil gerar ou mesmo utilizar o conhecimento. A informação é o lastro do conhecimento, expressando a manifestação do conhecimento explícito, sempre menor que o tácito, porém importantíssimo uma vez que permite

registrar o que é possível e ajudar resgatar a experiência individual e coletiva no interior das organizações. Informação não é sinônimo de conhecimento, mas está profundamente a ele atrelada. Gerar, armazenar e saber utilizar informação é grande diferencial competitivo, como o é criar ambientes no quais as pessoas se sintam estimuladas a gerar, compartilhar e utilizar o conhecimento para o próprio crescimento, de suas organizações e da própria sociedade.

5 RECUPERAÇÃO DA INFORMAÇÃO

5.1 Considerações Preliminares

Dados e informações, conceitos já discutidos anteriormente e não exaustivamente nesse trabaho, são realidades intrinsicamente ligados à vida e à trajetória da humanidade. Eles estão vinculados profundamente ao conhecimento gerado pelo homem. O ser humano tem consciência daquilo que é e do que pode vir a ser. Tem capacidade de refletir, de elaborar, de construir (e muitas vezes destruir) sobre a realidade em que vive. Não apenas caminha sobre a terra, mas procura entender o sentido dessa caminhada, fazendo-se ser histórico e criando sentido para si, no tempo e no espaço. Faz isso através da busca do entendimento do mundo e de si. Em sua relação com o mundo, ao longo de sua existência sobre o planeta terra, portanto, adquire conhecimento sobre si e sobre o mundo. Portanto, o homem busca de conhecimento para existir sobre o mundo e para alcançar esse conhecimento necessita obter informações, interagindo com outros seres humanos e objetos. Não só adquire conhecimentos, mas também os gera, compartilhando-os (muitas vezes) com outros homens. De certa forma, isso vem se estabelecendo em nosso jeito de ser e de viver, em nossa cultura. Por isso, ao longo dos séculos, o ser humano sedimentou expressiva concentração de itens ou objetos capazes de transmitir informação. Bibliotecas, museus, acervos de obras de arte, coleções particulares de indivíduos ou grupos foram sendo instituídas ao longo do tempo, bem como os esforços e processos para administrar e tratar esses agrupamentos de objetos de informação, aos quais podemos chamar de unidades ou centros de informação. A biblioteca obviamente destaca-se entre eles. Instrumentos de suporte foram desenvolvidos para que os ambientes propostos pelos centros de informação se tornassem gerenciáveis e pudessem ter continuidade. Esse processo demandou o estabelecimento da grande maioria dos métodos de armazenamento, localização e manipulação de informações que existem atualmente. Podemos assumir que, na complexidade do contexto sócio-cultural e econômico que caracteriza hoje nossa sociedade, recuperar, tratar e transformar a informação em substrato do conhecimento é alavanca para a competência e certamente hoje um dos maiores desafios organizacionais. O conhecimento é uma característica diferenciadora dos seres humanos e sua construção e utilização demanda a busca de informações.

Historicamente, de acordo com BUCKLAND (1997), por certo período de tempo, a atividade de produzir e gerenciar a literatura em geral foi chamada de bibliografia. A partir da revolução cientifica e industrial, no século XVII, somado ao desenvolvimento da imprensa, houve um crescimento rápido de publicações, o que demandou novos processos de organização de acervos para coleta, preservação, organização, representação, recuperação e reprodução de documentos. Surge então a documentação como um conjunto de técnicas requeridas pela gestão de documentos. Para esse autor, documentação tornou-se um termo mais genérico capaz de incluir a bibliografia, os serviços de informação, gestão de registros e atividade de arquivamento. Posteriormente, ao longo do tempo, especialmente a partir da década de 50, denominações mais adequadas passaram a ser usadas, tais como Ciência da Informação, Armazenamento da Informação ou Gerenciamento de Informações, ou então, simplesmente, Recuperação da Informação. A operacionalização das funções dessa área, tão importante para a Gestão do Conhecimento, se dá a partir da utilização de sistemas que cumprirão o papel de selecionar, inserir, descrever, representar, organizar, armazenar, recuperar e disponibilizar ao usuário os dados ou informações (dependendo da perspectiva teórica de um outro autor ou olhar do observador, uma vez que para certos autores, sistemas não lidam com informações, mas apenas com dados ou objetos potencialmente portadores de informação, como os documentos).

Nas últimas três décadas assistimos a um processo intenso de "virtualização" dos documentos associado também ao grande crescimento de suas bases, ou seja: cada vez mais os documentos se apresentam em grandes volumes, em muitas bases e sob formato digital, demandando cada vez mais sistemas informatizados para lidar com esses mesmos documentos.

5.2 O Conceito de Sistema

A idéia de sistema aplicada ao fenômeno da geração, tratamento e uso da informação é adequada por embutir na sua essência que tal atividade possui uma complexidade e envolve diferentes elementos. Um sistema pode ser entendido como um conjunto de elementos relacionados entre si, organizados de acordo com determinada lógica ou certos princípios, formando um conjunto global, um todo ou uma unidade. Há uma articulação, uma coerência com certos princípios básicos, ou mesmo com certos

objetivos ou funções que esse sistema possa ter. BERTALANFFY (1968) foi um autor importante no campo do estudo dos sistemas. Em sua obra Teoria Geral dos Sistemas, criticou as visões reducionistas na ciência propondo um modelo ampliado, que pressupõe uma interação das diversas realidades existentes entre si. Propôs uma visão de mundo como uma hierarquia de sistemas em que as idéias de interação e de todo articulado são importantes. Embora tenha iniciado seus estudos na Biologia, seu trabalho influenciou muitas áreas, incluindo as sociais e as exatas, obviamente com algum viés funcionalista, o que não diminui sua importância.

De acordo com a Teoria Geral dos Sistemas, apresentada em BERTALANFFY (1968), um sistema possui: entrada, processamento, saída e retroalimentação (feedback), conforme a figura 10. Essa visão teve aderência muito grande à idéia de sistemas de informação.

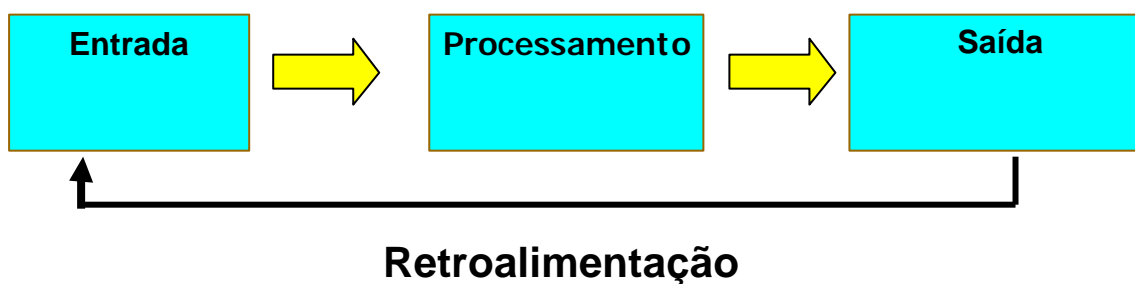



Figura 10: O conceito de sistema de BERTALANFFY
Fonte: Ilustração elaborada pelo autor

Aplicando tal idéia ao campo dos estudos da informação de selecionar, inserir, descrever, representar, organizar, armazenar, recuperar e disponibilizar ao usuário os dados ou informações chegaremos aos conceitos de sistema de informação e sistema de recuperação da informação. ARAÚJO (1994), baseando-se especialmente em PAO (1989) e SARACEVIC (1983), afirma que temos diferentes maneiras de subdividir um sistema que lida com informação e que uma delas é justamente a subdivisão nos três subsistemas abaixo descritos.

- Entrada – envolve os processos de seleção/aquisição, descrição, representação e organização de arquivos, bem como o armazenamento.
- Saída – engloba a questão da interface com o usuário com a análise e negociação de questões relacionadas ao comportamento de busca da informação, a partir do

entendimento das necessidades de informação dos usuários. São consideradas aqui as estratégias de busca/recuperação / e acesso aos dados/informações e as formas de entrega dos mesmos a tais usuários.

- Avaliação – processo que se refere tanto à entrada quanto à saída, procurando avaliar o custo/eficácia do sistema, os benefícios e o retorno por ele proporcionados, quão bem ele está funcionando, dentre outros aspectos.

Considerando a proposta teórica de BERTALANFFY (1968), algumas das atividades inseridas nos processos de entrada e de saída poderiam ser reclassificadas na "caixa" do "processamento" acima apresentada. Como exemplo, talvez pudessem assim ser a representação e questões relacionadas às estratégias que o sistema deve processar para organizar, tratar e recuperar dados/informação.

5.3 Sistemas de Informação e Sistemas de Recuperação da Informação

As expressões Sistemas de Informação e Sistemas de Recuperação da Informação são usadas às vezes como equivalentes às vezes de forma diferenciada. Inicialmente poderíamos ver na expressão "recuperação" uma idéia restritiva para o sistema, uma vez que, na prática, não temos só a idéia de recuperar, mas também de inserir, manter, armazenar, atualizar e organizar os objetos de informação mantidos pelo sistema. Entretanto, em muitos escritos e muitos autores, um exame mais atento permite definir tais expressões como equivalentes. Por outro lado, a idéia de "sistema de informação" na literatura recente embute uma visão mais generalista e um componente tecnológico mais forte, mormente influenciado pela Ciência da Computação. Já a idéia de sistemas de recuperação da informação embute uma visão mais específica, em que a idéia de documento, em geral textual é mais central. De qualquer forma, há variação no uso e no significado dos dois termos para diferentes autores.

Na perspectiva de TURBAN (2003), que tomaremos como referência nesse trabalho para a idéia de sistema de informação, tal sistema coleta, processa, armazena, analisa e dissemina informações com um propósito específico. Como qualquer outro sistema, um sistema de informação abrange entrada (dados), processamento (cálculos, organização e tratamento dos dados) e saídas (informações e relatórios). Portanto, é um conjunto de componentes relacionados, trabalhando juntos para coletar, recuperar,

processar, armazenar e distribuir informação. Também na perspectiva desse autor, nem todos os sistemas de informação são computadorizados. Um sistema de informação baseado em computador é aquele que usa o computador e a tecnologia das telecomunicações para executar suas tarefas. Uma tecnologia de informação é um componente particular de sistema (por exemplo, um computador pessoal, uma impressora ou uma rede). Poucas tecnologias de informação são usadas isoladamente. Os principais componentes de um sistema de informação baseado em computador são: o hardware, o software, as bases de dados, registros ou documentos (conforme o caso), o processo de conectividade que viabiliza o compartilhamento de recursos (rede), os procedimentos (estratégias, políticas e métodos para usar os sistema de informação) e as pessoas (elemento mais importante que inclui usuários finais e operadores do sistema). Isso pode ser representado na figura 11 abaixo:

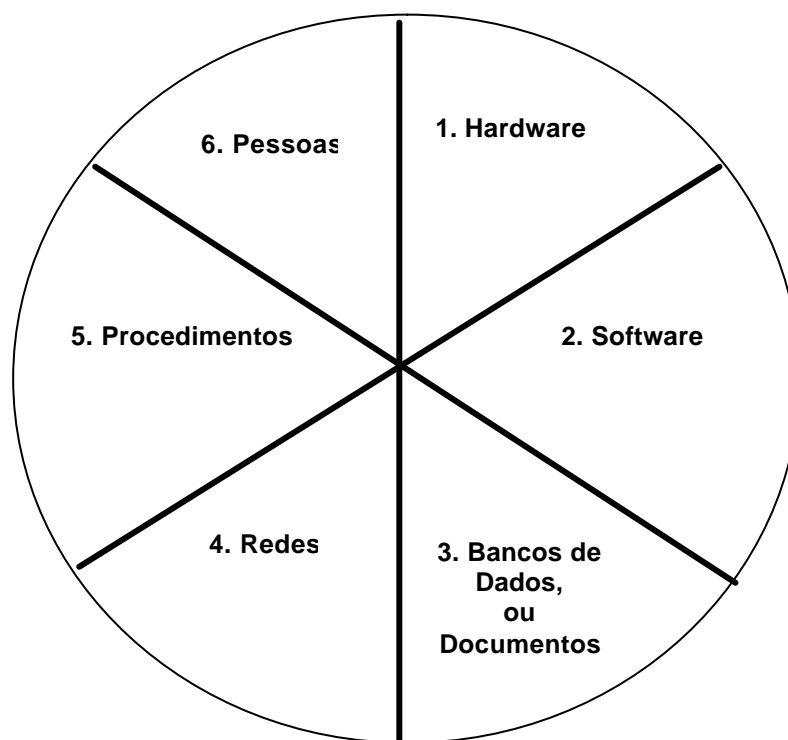



Figura 11 Componentes de um Sistema de Informação

Fonte: Adaptado de TURBAN (2003)

Um sistema de informação é, portanto, um sistema sócio-técnico, composto por um conjunto de componentes relacionados entre si, que coletam, processam, armazenam

e distribuem informação para públicos nela interessados. Como sistema sócio-técnico, é utilizado, concebido e operado por pessoas e possui hardware, software, bases de dados, recursos de conectividade (rede), procedimentos de uso do mesmo. A mineração de textos, objeto do presente trabalho, pressupõe o uso de um sistema de informação baseado em computador para viabilizar seu propósito, que é o de permitir a recuperação inteligente e automática de documentos digitais, servindo-se de recursos computacionais para atingir tais objetivos.

Para SALTON & MCGILL (1983) temos diferentes tipos de sistemas de informação, quais sejam:

- Sistemas de Recuperação da Informação
- Sistemas Gerenciadores de Banco de Dados
- Sistemas de Informações Gerenciais
- Sistemas de Suporte à Decisão
- Sistemas de Recuperação de Fatos (*query answering systems*)

Há uma grande variedade e diferentes possibilidades de classificação de sistemas que lidam com a informação. VICKERY & VICKERY (1987), citado por ARAUJO (1994) enumeram oito critérios principais para diferenciação e criação de uma tipologia dos sistemas que lidam com informação. De acordo essa visão, podem ser utilizados os seguintes pontos de discernimento:

- em relação ao processo de transferência da informação, que implica em troca de mensagens, estas podem ser registradas (em base de dados, ou qualquer documento) ou não registradas;
- em relação às audiências do usuários, podemos tê-los concentrados ou dispersos;
- em relação à natureza da audiência, o número de usuários que podem ser atingidos simultaneamente podem ser individual, uma comunidade restrita ou destinado à uma massa de usuários;
- em relação à coleta de informações, as fontes podem ser também concentradas e dispersas;

- em relação à sua forma de operação, o sistema pode operar de forma centralizada ou descentralizada (em rede), com a definição de pontos de aquisição em que as informações são coletadas por ele, estabelecimento das estruturas de armazenamento e pontos de acesso em que a informação é fornecida aos usuários;
- em relação ao tipo de usuário, o que pode variar muito, indo desde um tomador de decisão em uma organização até um jovem que busca noticias esportivas na *Internet*;
- em relação ao meio utilizado para entregar as mensagens aos usuários, podendo ser de forma oral, manuscrita ou eletrônica;
- em relação à forma da informação, o que permite classificar os sistemas como lidando com dados e fatos, documentos, referência a documentos, fontes de referência a pessoas e instituições, etc.

Podemos adicionar uma nona característica que seria sobre a natureza constitutiva ou estrutura do próprio sistema, em que poderíamos classificá-lo como *informatizado* ou *não informatizado*, e uma última característica distintiva, baseando-se em VICKERY & VICKERY (1987), citado por ARAUJO (1994), que é a sua finalidade ou propósito, ou seja: para qual função tal sistema foi concebido e desenvolvido, o que pode ser para controle de um acervo, fornecimento de informação gerencial a executivos tomadores de decisão, informação científica a pesquisadores, ou informações operacionais a pessoas que trabalham no *chão-de-fábrica* em uma indústria. Isso significa que, sob o ponto de vista organizacional, então que temos a classificação de sistemas de informação estratégica ou gerencial e sistemas de informação transacional ou operacional.

A *"forma da informação"* ou *"dos dados"*, dependendo da perspectiva, tem sido talvez o critério mais forte utilizado para a classificação dos sistemas. Os sistemas podem lidar com registros em bancos de dados relacionais, textos, imagens, sons e outros itens multimídia. Na segunda metade do século XX duas fortes linhas se consolidaram: a linha dos sistemas que tratam dados e dos sistemas que tratam documentos. A linha do tratamento de dados cresceu vertiginosamente em função da ampla proliferação de sistemas de informações em empresas para processamento de

transações de negócios. Já a linha do processamento de documentos evoluiu na trajetória da área da Biblioteconomia e da Gestão de Documentos no que se afirmou sob a denominação de Recuperação da Informação. Podemos dizer que tais vertentes geraram mesmo dois diferentes paradigmas: o da área de Banco de Dados – fortemente vinculado à idéia de dados relacionais manipulados pelos DBMS (*Database management systems*, ou sistemas gerenciadores de banco de dados) e o da área de Recuperação da Informação. Ao longo dos anos 90, o critério do tipo de usuário foi muito salientado, com uma forte discussão sobre a separação dos sistemas em gerenciais (ou estratégicos) e transacionais (operacionais). Tais classificações não são mutuamente excludentes, mas sim visões por pontos de vista diferentes de um mesmo fenômeno: os sistemas que lidam com o desafio de transformar dados em informações para um usuário ou uma comunidade usuários. Sob os pontos de vista, conceitual, metodológico e tecnológico, considerar a forma dos dados é um fator importante a ser avaliado, uma vez que essa forma determinará o esforço computacional para tratá-lo.

De acordo com SOUZA (2006):

> *"há que se distinguirem os sistemas de recuperação de informações (SRI) dos sistemas de gestão de bancos de dados (SGBD). Dados podem ser definidos como seqüências de símbolos para os quais são atribuídos significados; símbolos estes que podem ser codificados, interpretados e manipulados por programas de computador, e enviados através de redes e dispositivos de comunicação. O conceito de informação já carrega um grau maior de abstração. A informação não prescinde do sujeito que a depreenda a partir dos dados, no ato conhecido como interpretação. No sentido estrito do conceito, nenhum programa de computador lida, sob o ponto de vista da máquina, com informações, a não ser que possua alguma capacidade de arrazoamento, e, assim mesmo, a utilização do termo dá margem a discussões. No uso corrente, porém, ambos os termos são utilizados para sistemas, apesar das diferenças entre os sistemas de recuperação de informações e sistemas de recuperação de dados, como os SGBDs" (p.163).*

É realmente importante separarmos os dois paradigmas. Em um SGBD, que lida com tabelas no modelo relacional, uma consulta, formulada em uma linguagem de manipulação de dados, possui exclusivamente uma só resposta possível e correta, isto é, existe ou não existe um conjunto resposta de elementos que atendem exatamente ao que o usuário está procurando ou não. Por outro lado, para os chamados SRIs, há sempre um grau de incerteza associada ao tipo de documento e, por isso, é possível que não exista uma única resposta exata à consulta do usuário. Na perspectiva de FRAKES (1992), tais

sistemas são difusos e probabilisticos, não trabalhando com a exatidão dos SGBDs, sendo que um SRI em geral recupera uma aproximação, apresentando diferentes respostas possíveis e construindo um ranking em que os documentos são enumerados sequencialmente de acordo com uma estimativa de relevância.

BAEZA-YATES & RIBEIRO-NETO (1999) afirmam que recuperar dados no contexto de um SRI consiste principalmente em determinar quais documentos de uma coleção contém a palavra chave na consulta de um usuário embora, muito frenqüentemente, isso não seja suficiente para satisfazer a necessidade de informação dele. Na verdade, o usuário de um SRI está mais preocupado em recuperar informações sobre um assunto do que com recuperação de dados que satisfaça a uma determinada consulta. Para esses autores, uma linguagem de recuperação de dados objetiva recuperar todos os objetos que satisfaçam a condições claramente definidas tal como aqueles em uma expressão regular ou em uma expressão de álgebra relacional. Assim sendo, para sistemas de recuperação de dados, um simples objeto errado dentro de um milhão de objetos recuperados é uma falha total. Para um SRI, entretanto, objetos recuperados com pequenos erros provavelmente passariam despercebidos. A principal causa dessa diferença é que a recuperação de informações usualmente lida com textos em linguagem natural, os quais nem sempre são bem estruturados e podem ter ambigüidades semânticas. Por outro lado, sistemas de recuperação de dados, tais como os bancos de dados relacionais lidam com dados bem definidos em estrutura e semântica.

Continuando na perspectiva de análise BAEZA-YATES & RIBEIRO-NETO (1999), temos que a recuperação de dados, enquanto proporciona uma solução para os usuários de sistemas de bancos de dados não resolve o problema da recuperação de informação sobre um sujeito ou tópico. Em sua tentativa de efetivamente satisfazer as necessidades de informação dos usuários, sistemas de recuperação da informação buscam de alguma maneira interpretar os conteúdos de itens de informação em uma coleção, organizando-os e escalonando-os de acordo com o nível de relevância para a consulta de um usuário. Tal interpretação de conteúdo implica na extração de informações sintáticas e semânticas dos textos dos documentos e no uso dessas informações para casamento com as necessidades de informação do usuário. É difícil não somente conhecer como extrair essas informações sintáticas e semânticas, mas também como usá-las para decidir sua relevância.

Segundo ABITEBOUL, BUNEMAN et alli. (2000) os dados se apresentam em três formas distintas: *estruturados*, *semi-estruturados* e *não estruturados*. Analisando sua proposta, podemos caracterizar tais formatos da seguinte maneira:

- Dados estruturados – aqueles que apresentam uma representação estrutural homogênea, com seu tipo bem definido. Sua forma de representação é bem clara, feita de forma documentada, através de descrição de *lay-out* com base na documentação do aplicativo. Nestes dados inclui-se o cadastro de funcionários, clientes, produtos e as respectivas tabelas dos bancos relacionais que lhes dão sustentação, dentre outros.

- Dados semi-estruturados – aqueles que se apresentam com uma representação estrutural heterogênea, não são completamente estruturados e nem tem seu tipo totalmente definido. Entretanto, possuem alguma estrutura. Os dados publicados na *Web* são bons exemplos. Em alguns casos os dados portam uma descrição uniforme como catálogos de produtos, em outros casos os dados possuem formato que podem ser identificados como textos com alguma estrutura, contratos e relatórios que possuem campos ou seções relativamente bem definidas e identificáveis. Neste tipo de dado a descrição do formato está, em geral, implícito ao documento.

- Dados não estruturados – são aqueles em que é impossível ou muito difícil encontrar a forma de representação. Enquadram-se neste tipo de dados textos diversos e imagens, por exemplo. Em tal caso, não há geralmente informação descritiva associada.

Podemos assumir que os dados estruturados tem sido, ao longo do tempo, o foco de atenção do paradigma da área de banco de dados, enquanto os dados semi-estruturados e os não estruturados os da área de recuperação da informação. Impulsionada pela demanda gerada pela *Web*, a área de recuperação da informação, têm sido objeto de várias pesquisas a fim de fornecer suporte ao atendimento das necessidades de informação das organizações. O volume de documentos disponíveis

tem crescido consideravelmente e a tarefa de filtrar informação relevante passou a representar uma boa dose de tempo para as organizações.

5.4 O Paradigma da Área de Banco de Dados e Sistemas Vinculados

Um aspecto muito importante para a tecnologia da informação foi o surgimento e crescente utilização de bancos de dados e de seus sistemas gerenciadores de bancos de dados - os SGBD's, que consistem em uma coleção de dados inter-relacionados e um em um conjunto de programas para acessá-los. O conceito de dado já foi discutido no capítulo anterior. A partir dele podemos refletir sobre a definição de processos tecnológicos de organização e processamento de dados e de sua importância para o desenvolvimento dos sistemas de informação. Do ponto de vista tecnológico vai do menor nível (bit – digito binário), passando por bytes, campos, registros e arquivos até chegar ao banco de dados. Foi, portanto, necessário estruturar e hierarquizar dados para se obter resultados interessantes no que tange à geração de informação.

Antes do advento dos chamados sistemas gerenciados de bancos de dados – SGBDs a área de informática trabalhava com o chamado processamento de arquivos em que os dados eram armazenados em estruturas de arquivos pouco flexíveis e muito seqüenciais. Porém, havia problemas de inconsistência, isolamento e redundância dos dados, dificuldade de acesso a eles, problemas de integridade, atomicidade, dificuldades para o acesso por usuários ou aplicações simultâneas e segurança. Os SGBDs surgiram como uma busca de resposta eficaz a tais problemas. SGBDs são conjuntos de dados associados a um conjunto de programas para acessá-los, conforme a visão de KORTH, SILBERSCHATZ & SUDARSHAN (1999). O conjunto de dados, comumente chamados de bancos de dados, contém informações sobre uma organização ou uma realidade em particular.

Os SGBDs tratam os conjuntos de dados e operam tais programas com uma estrutura lógica que permite melhor organizar o processo de geração da informação a partir de modelos conceitual, lógico e físico de estruturação de dados nas bases dos

sistemas aplicativos[37] que utilizam as bases de dados por eles requeridas. SOUZA (2006) comenta que:

> *"Em sistemas gerenciadores de bancos de dados, os símbolos são armazenados em uma estrutura matricial em campos determinados, com metadados que lhes conferem certo sentido ontológico. Para recuperar dados específicos, basta especificar as restrições necessárias aos campos de pesquisa e codificá-las numa questão ou query (argumento de entrada no sistema) para que se tenha a resposta exata, fruto de busca completa e exaustiva"*(p.163).

Já KORTH, SILBERSCHATZ & SUDARSHAN (1999), analisam que os sistemas de bancos de dados são projetados para gerenciar grupos de dados que serão a matéria prima da informação, pelo menos no âmbito dos sistemas nele baseados. Através das linguagens de manipulação de dados eles permitem inserir, recuperar, remover ou modificar os registros de dados armazenados nos bancos de dados. No nível físico, devem operar com algoritmos eficientes para executar tais funções e no nível abstrato, da interface com o usuário, deve proporcionar uma interação eficiente entre desses com o sistema. Os SGBDs fornecem segurança contra falhas, evitam redundância e inconsistência, garantem um armazenamento mais adequado, melhoram o acesso aos dados e solucionam problemas de integridade.[38] O uso de bancos de dados cresceu desde o advento da tecnologia da informação, uma vez que as estruturas de dados por eles armazenadas podem ser utilizadas várias vezes em muitos sistemas diferentes, e isso ficou ainda mais ressaltado com a utilização crescente do modelo relacional de dados. Os sistemas gerenciadores de banco de dados, conforme TURBAN (2003) possuem quatro componentes:

- o modelo de dados, que define o modo como os dados são organizados e estruturados em termos conceituais. Por exemplo, podem ser organizados segundo o modelo, hierárquico, em rede, relacional, orientado a objetos, objeto-relacional, hipermídia e multidimensional;
- uma linguagem de definição de dados, que define os tipos de informações existem no banco de dados e como estão estruturadas;

---

[37] Um sistema aplicativo é qualquer programa de computador que tem uma finalidade ou aplicação bem especifica para uma necessidade devidamente delimitada para certos grupos de usuários. Por exemplo, um software comercial que controla vendas e faturamento ou um programa para controle contábil, são bons exemplos.

[38] Para análise mais ampla, consultar estudo detalhado na obra KORTH, H., SILBERSCHATZ, A. & SUDARSHAN, S Sistema de Banco de Dados. 3ª. Edição Brasileira. São Paulo. Makron Books, 1999.

- uma linguagem de manipulação de dados, que define as formas de manipulação dos dados nas bases de dados;
- um dicionário de dados, que armazena definições de elementos dos dados e suas características. Um elemento de dado representa um campo. Padroniza e cria dados sobre os dados (metadados).

Abordaremos a seguir o modelo relacional por sua representatividade computacional e comercial e pelos efeitos significativos que produziu na informática comercial e na sociedade cada vez mais computadorizada.

5.4.1 A Perspectiva do Modelo Relacional

Não é propósito de que esse trabalho discuta de forma ampla esse assunto tão complexo, mas ele pretende apenas pontuar uma reflexão resumida e útil para o contexto da presente tese. Podemos postular que o chamado modelo relacional criou não só uma forma de lidar com quantidades crescentes de dados, mas significou também a solidificação de um paradigma que podemos denominar o do processamento e recuperação de dados estruturados de forma relacional. Ao longo da década de 60, os computadores começaram a ser utilizados nos negócios, e sistemas de informação e linguagens de programação foram desenvolvidos especificamente para o uso corporativo. Mas havia, até então, apenas dois modelos de organização de dados: o modelo hierárquico e o modelo em rede[39], ambos com limitações para lidar com bases de dados crescentes, embora buscando maior estruturação. Foi o trabalho pioneiro de Edgar Codd, de proposição do chamado “Modelo Relacional”, que apresentou uma opção mais simples, que não exigia de quem estivesse acessando o sistema um profundo conhecimento de programação. É ponto capital a publicação do artigo de Codd "*O Modelo Relacional de Dados para Grandes Bancos Compartilhados*" (*A Relational Model of Data for Large Shared Data Banks*), em 1970. O modelo propôs o uso da teoria matemática de relações (álgebra relacional). É baseado na lógica de predicados e teoria dos conjuntos. O postulado principal do modelo relacional é o de os dados podem ser representados por relações matemáticas ou um subconjunto do produto cartesiano de vários conjuntos.

[39] O modelo hierárquico tratava os dados sob a perspectiva de uma estrutura de árvore de instâncias sucessivamente subordinadas, enquanto o modelo em rede podia permitir algumas relações lógicas mais complexas entre representações de dados.

O modelo redundou em uma grande transformação no campo da informática comercial, uma vez que propiciava aos bancos de dados uma organização muito diferente dos dois sistemas usados até então. No modelo hierárquico, todos os registros eram relacionados com outro, acima na hierarquia (um registro principal encabeça a lista, e todos os outros dados vinham abaixo) e, no modelo de rede relacionada uma informação poderia estar listada em duas hierarquias distintas. Diferentemente, no modelo relacional, CODD (1970) propôs que os dados fossem armazenados em tabelas simples, constituídas por linhas e colunas e que tais dados fossem relacionados entre si. Conforme analisa KORTH, SILBERSCHATZ & SUDARSHAN (1999):

> *"um banco de dados relacional consiste em uma coleção de tabelas, cada uma das quais um nome único. Uma linha em uma tabela representa um relacionamento entre um conjunto de valores. Uma vez que essa tabela é uma colação de tais relacionamentos, há uma correspondência estreita entre o conceito de tabela e o conceito matemático de relação, a partir das quais se origina o nome desse modelo de dados"*(p.61).

Portanto, o modelo relacional inclui uma coleção de estruturas de dados, denominada formalmente de relações ou de modo informal de tabelas. Além disso, operações matemáticas deveriam ser usadas para recuperá-las destas tabelas (tais como *UNION*, *SELECT*, *SUM* etc.). Isso redundou no estabelecimento de uma coleção dos operadores, da álgebra e do cálculo numa perspectiva relacional os quais passaram a constituir o fundamento da linguagem *SQL – Structured Query Language*.

Três aspectos básicos foram alcançados com o modelo relacional:

a) uma maior independência dos dados, com definição mais explícita dos limites entre os aspectos físicos e lógicos de um gerenciador de banco de dados;

b) permitir um modelo estrutural simples de forma que usuários de várias categorias pudessem ter um entendimento comum dos dados e pudessem se comunicar através do banco de dados;

c) uso de linguagem de alto nível - existência de uma linguagem que permitisse a manipulação de um conjunto de dados através de apenas um simples comando – os previstos pela linguagem *SQL*.

Os bancos de dados relacionais passaram a permitir aos projetistas, programadores e usuários escreverem consultas (*queries*), reorganizando e utilizando os dados de modo flexível e não necessariamente do modo definido pelos projetistas originais. Tal flexibilidade se tornou especialmente importante em bases de dados que deviam ser utilizadas por muito tempo, tornando as bases de dados relacionais muito importantes no meio comercial. Por isso se difundiu intensamente, sendo até hoje ainda a principal referência em modelos de dados com aplicações organizacionais. Deflagrou uma série de desenvolvimento de sistemas de informação aplicativos à diferentes necessidades das organizações e continuam vigorosos nos dias de hoje. No modelo relacional, o SGBD administra a integração de tabelas de dados entre si que podem ter os mais diversos arranjos. Por isso, tornou-se muito adequado para suportar as aplicações (sistemas) específicos das organizações, desde as de pequeno porte, até os sistemas de gestão integrada, os chamados *ERP – Enterprise Resource Planning*. Com as pressões por aumento de competitividade, nos anos 80 e 90, grande parte das empresas buscaram fortemente automatizar ao máximo suas operações, o que levou a um forte desenvolvimento de sistemas transacionais – voltados para as transações do dia-a-dia dos seus negócios e os bancos de dados relacionais tiveram e ainda mantêm papel importante no que tange a isto (figura abaixo).

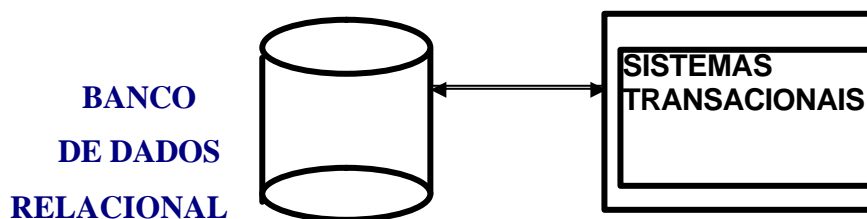



Figura 12: Os bancos de dados armazenam dados e alimentam os sistemas de informação.

Em suma, a utilização da idéias de CODD (1970) levou a um paradigma não só de gerenciamento de bases de dados, como também de diferentes tipos de sistemas baseados em sua concepção. As linguagens de manipulação de dados, especialmente a linguagem *SQL*, que passou a ter inclusive padrões aceitos pela indústria de software,

foram intensamente difundidas e é, ainda hoje, referência em definição e manipulação de dados atualmente, sobretudo em grandes aplicações. Os sistemas baseados na visão relacional não só utilizam como geram grande quantidade de dados relacionais, registros relativos às transações – operações de uma organização. Essa base de dados torna-se substrato para geração de informações a múltiplos e diferentes usuários.

5.5 O Paradigma da Recuperação de Informação

5.5.1 Visão Geral

Uma grande quantidade de informação pode hoje ser produzida e disponibilizada com facilidade, especialmente após a difusão do uso do computador. Com a elevação da quantidade de documentos, o processo de recuperação de informação não é uma tarefa simples. De acordo com MEADOWS (1992) recuperar informação é encontrar a informação que se deseja em uma base de dados ou armazém de informação. Podemos ampliar a análise de tal definição afirmando que recuperar informação é encontrar a informação desejada em uma coleção de objetos capazes de transmitir informação, sendo que essa coleção pode ser uma base de dados, uma base de documentos textuais, uma biblioteca, uma pinacoteca, um museu, um acervo de imagens e sons, etc.

Recuperar informações significa recuperar objetos de informação (ou seja, documentos) que possivelmente contenham informações para o usuário (LANCASTER 2004). Um documento, conforme discutido no capitulo anterior, é um objeto do mundo real com potencial para transmitir informação a alguém. Ao ser recuperado, em um contexto de busca, é porquê faz algum sentido no contexto de um usuário específico.

Diferentemente dos dados manipulados por sistemas gerenciadores de bancos de dados relacionais, e anteriormente a eles, dados expressos sob a forma de documentos textuais, sejam eles eletrônicos ou não, já colocavam desafios ao esforço de buscar respostas para as necessidades de informação humanas. Não obstante existam outros eventos anteriores a considerar, é a partir do fenômeno da “explosão da informação”, logo após a II Guerra Mundial, que a área de Recuperação da Informação começa a ganhar maior densidade e identidade. Os acervos de registros ou documentos crescem e precisam ser alvo de processos de tratamento e recuperação para serem efetivamente

úteis aos usuários. BUSH (1945), em seu célebre artigo já citado, propõe o uso de novas tecnologias da informação para viabilizar o acesso de modo satisfatório, embora tais tecnologias ainda estivessem em um grau de desenvolvimento principiante. Posterior e gradativamente os documentos eletrônicos (digitalizados) não pararam mais de crescer em volume. A estrutura que possuem é distinta das tabelas manipuladas pelos SGBDs. Precisam de estratégias adequadas de manipulação e recuperação. Portanto, o foco central da discussão a seguir é sobre documentos textuais.

A expressão "Recuperação de Informações" foi proposta por MOORES (1951), um empresário atuante na área e que definiu "recuperação de informações" como o

> "*nome do processo ou método onde um posssível usuário de informação pode converter a sua necessidade de informação numa lista real de citações de documentos armazenados que contenham informações úteis a ele (...) recuperação de informação abarca os aspectos intelectuais da descrição da informação e a sua especificação para busca, assim como também quaisquer sistemas, técnicas ou máquinas que sejam empregadas para efetuar a operação* (p.20-32,)"

Calvin Moores procurava encontrar respostas para a necessidade de armazenar e gerenciar eficazmente a informação, e conforme ARAÚJO (1994), permitir o acesso rápido e preciso ao crescente volume de documentos que passava por exponencial crescimento desde o século XVII e pelo surgimento do computador, o qual passou a ser visto como potencial solução para os problemas de recuperação da informação. Para OLIVEIRA (1) (2005), a concepção proposta por MOORES (1951), implica em três questões importantes:

- Como descrever intelectualmente a informação?
- Como especificar intelectualmente a busca?
- Quais sistemas, técnicas ou máquinas devem ser empregados?

Ainda segundo essa autora,

> *"as atividades desenvolvidas no âmbito da temática "recuperação da informação" conduziram a estudos teóricos e conceituais sobre a natureza da informação; a natureza dos seus registros (incluindo a bibliometria); os estudos relativos ao uso e aos usuários da informação; estudos do comportamento humano frente à informação; a interação homem computador,*

dentre outros. Enfim, a recuperação da informação possibilitou o surgimento dos sistemas automatizados de informação" . (OLIVEIRA (1), 2005: 12).

O processo de recuperação da informação envolve três elementos importantes: o usuário, o sistema de recuperação e a base de documentos. No processo de recuperação da informação, o objetivo maior de um sistema de recuperação automatizado é fazer com que o usuário encontre a informação de que necessita rapidamente, de modo que esse usuário não precise analisar, ele próprio, as informações existentes na base de documentos.

Conforme já citado anteriormente, é importante salientar que a conceituação do que sejam os termos: dados, informação ou conhecimento é polêmica e não há a intenção do presente trabalho em explorar tais divergências e exaurir os múltiplos conceitos e as múltiplas divisões colocadas por diversos autores. Algumas questões de ordem, entretanto, são importantes. Primeiro: podemos avaliar a pertinência ou não da expressão "Recuperação da Informação". Se assumirmos, conforme ARAÚJO (1994) que a informação é algo dinâmico, que faz com que o agente modifique seu estado de conhecimento do mundo atual, modificando estruturas, e só tendo sentido no contexto da ação, "Recuperação da Informação" pode ser um termo inadequado. Além disso, a informação depende do sujeito, do usuário para conferir significado aos dados que foram recuperados para responder a uma determinada necessidade.

Baseando em (FOSKETT, 1997, p. 5), SOUZA (2006) afirma que:

"A recuperação de informações traz dificuldades intrínsecas ao conceito de "informação", como a dificuldade da determinação da real necessidade do usuário e do seu melhor atendimento com os documentos que fazem parte do acervo do sistema A associação entre os registros e seus conteúdos informativos é vaga, e isso pode acarretar problemas nas respostas a questões específicas, como baixas taxas de revocação e precisão".

Entretanto, a expressão "Recuperação da Informação" se consolidou na perspectiva da definição formulada por MOORES, acima citada e a assumiremos no contexto desse trabalho. De acordo com BAEZA-YATES & RIBEIRO-NETO (1999), as bibliotecas figuram entre os primeiros tipos de instituição a adotar SRI. A primeira

geração de SRI se limitava a automatizar tecnologias já utilizadas, tais como o uso de catálogos. Em uma segunda geração, foram acrescentadas funções de busca, pelo uso das palavras-chaves e outros tipos de consultas mais complexas. Na terceira geração, que se encontra em curso, o principal foco está em criar interfaces gráficas, formulários eletrônicos, características de hipertexto, e arquiteturas de sistemas abertos. A *Web* e as Bibliotecas Digitais configuraram um cenário novo para a área de Recuperação da Informação. Se considerarmos as máquinas de busca da web atualmente, concluímos que elas continuam usando índices similares aos utilizados por bibliotecários há séculos atrás. Entretanto, *t*rês mudanças drásticas ocorreram durante o avanço da tecnologia computacional e no crescimento da *web*, quais sejam:

- Atualmente há custos menores em ter acesso a várias fontes de informação. Isso permite que seja realizado um número de pesquisas muito grande, algo talvez impensável anteriormente.

- O desenvolvimento de todos os tipos de comunicação digital produziu um acesso ainda maior às redes. Isso implica que as fontes de informações estão disponíveis, mesmo que localmente distantes e que o acesso pode ser realizado rapidamente.

- A liberdade de publicar e divulgar qualquer tipo de informação que uma pessoa considere útil ou importante para ela. Isso aumenta cada vez mais a popularidade da *web*. Foi inaugurado um novo momento na história, no qual muitas pessoas têm acesso livre a uma enorme quantidade de publicações de médio e pequeno porte. Basicamente a sociedade passou a contar com baixo custo, maior acesso, e liberdade para publicações, o que têm permitido às pessoas usarem a *web* e as bibliotecas digitais modernas como a forma mais interativa de comunicação.

Esses aspectos conferem à Área de Recuperação da Informação particular importância para a sociedade nos dias atuais.

5.5.2 Conceito e Processos dos Sistemas de Recuperação da Informação

Como ponto de dessa discussão faz-se mister pontuar a importante contribuição de Wilfrid Lancaster, professor da *Graduate School of Library and Information Science* da Universidade de Illinois (EUA) que é um importante autor do campo da Recuperação da Informação. De acordo com LANCASTER (1968), conforme o tipo de aplicação que se dá aos Sistemas de Recuperação da Informação – SRIs, eles podem ser classificados em sistemas de recuperação de documentos, sistemas de recuperação de referencias, sistemas de recuperação de dados e sistemas de recuperação de fatos (ou "*query answering systems*"). Ainda de acordo com ele, os SRIs não informam o usuário no sentido de mudar seu conhecimento sobre objeto de sua questão, mas apenas o informa sobre a possível existência de documentos atinentes à questão, além de características desses documentos. Posteriormente, LANCASTER & WARNER (1993) os SRIs promovem o elo entre uma coleção de recursos de informação, em meio impresso ou não, e um grupo de usuários. Eles cumprem as tarefas de aquisição, armazenamento, organização e controle de documentos, e sua respectiva entrega e disseminação aos usuários. Ressaltando, LANCASTER (2004) postula que recuperar informações, no contexto de SRI, significa recuperar documentos que (supostamente) contenham informações relevantes para o usuário. Analisando os estudos desse autor e cotejando-os com ARAÚJO (1994), que propõe a idéia de que um Sistema de Recuperação da Informação possui os subsistemas de entrada, saída e avaliação, podemos construir uma representação.

Um SRI pode ser visto como uma moeda que possui duas faces, dois subsistemas ou, dependendo do ponto de vista, dois sistemas acoplados em torno de um núcleo comum: a base de documentos. Por um lado, temos a face da gestão documental e por outro a face da gestão das relações com o usuário. Isso pode ser representado com a figura 13. Na face da gestão documental, temos a entrada, processamento e saída dos documentos com potencial para responder às necessidades de informação do usuário, envolvendo as atividades de aquisição, inserção, representação (indexação), organização e armazenamento dos documentos. Na gestão documental temos como centro da atenção a descrição e representação dos documentos e das informações neles contidas, a partir das características de modelagem do próprio SRI. Na face da gestão das relações

com o usuário, temos a entrada, processamento e saída das requisições e interações dos usuários para que possam responder às suas necessidades de informação. Os usuários devem expressar suas necessidades de informação através de alguma forma de consulta ou navegação que seja compatível com o sistema e ele seja capaz de responder. Isso implica em atividades de definição de estratégias de busca, formulação, submissão e processamento de consultas, navegações e interações diversas do usuário com o SRI. A partir desses dois aspectos, o SRI deve viabilizar o casamento que vincule a consulta do usuário e as informações do sistema, para especificar quais informações são relevantes.

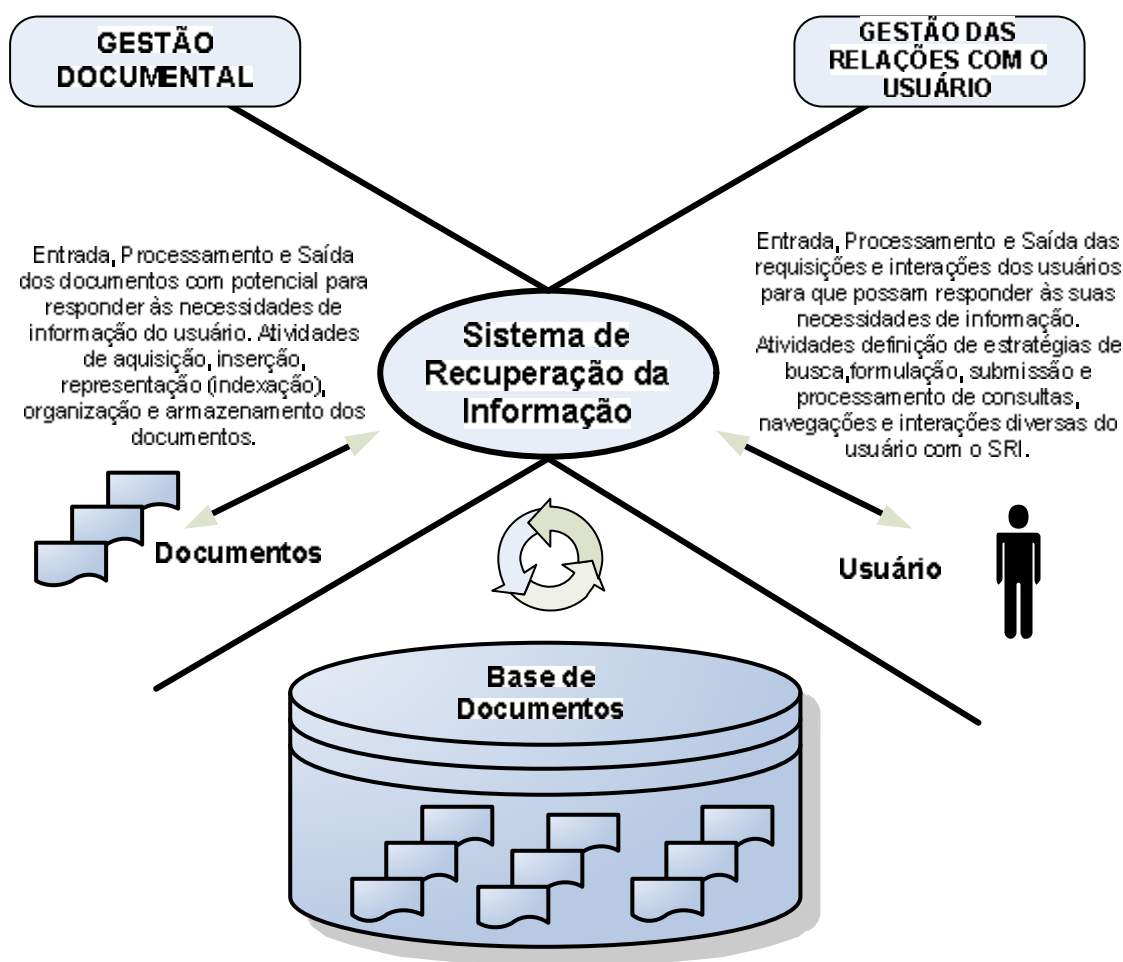



Figura 13: As Duas Faces de um Sistema de Recuperação da Informação
Fonte: Ilustração elaborada pelo autor

Detalhando essa visão, podemos recorrer a CENDON (2005) que afirma que:

> *"normalmente os sistemas de recuperação da informação lidam com documentos que contem principalmente texto e esse é seu grande desafio, já que devem se defrontar com a ambigüidade da palavra na recuperação da*

> *informação. (...) Em uma definição mais ampla, podem, também lidar apenas com a informação textual. (...) Um sistema de armazenamento e recuperação de desenhos de engenharia, por exemplo, seria também um SRI. Já um catálogo de fichas seria um exemplo de SRI não automatizado. (p.62)"* .

A autora lembra que tais sistemas tendem a ser cada vez mais automatizados e que automação oferece mais vantagens e possibilidades e seu uso vem aumentando. Também em suas palavras, analisando essa automação:

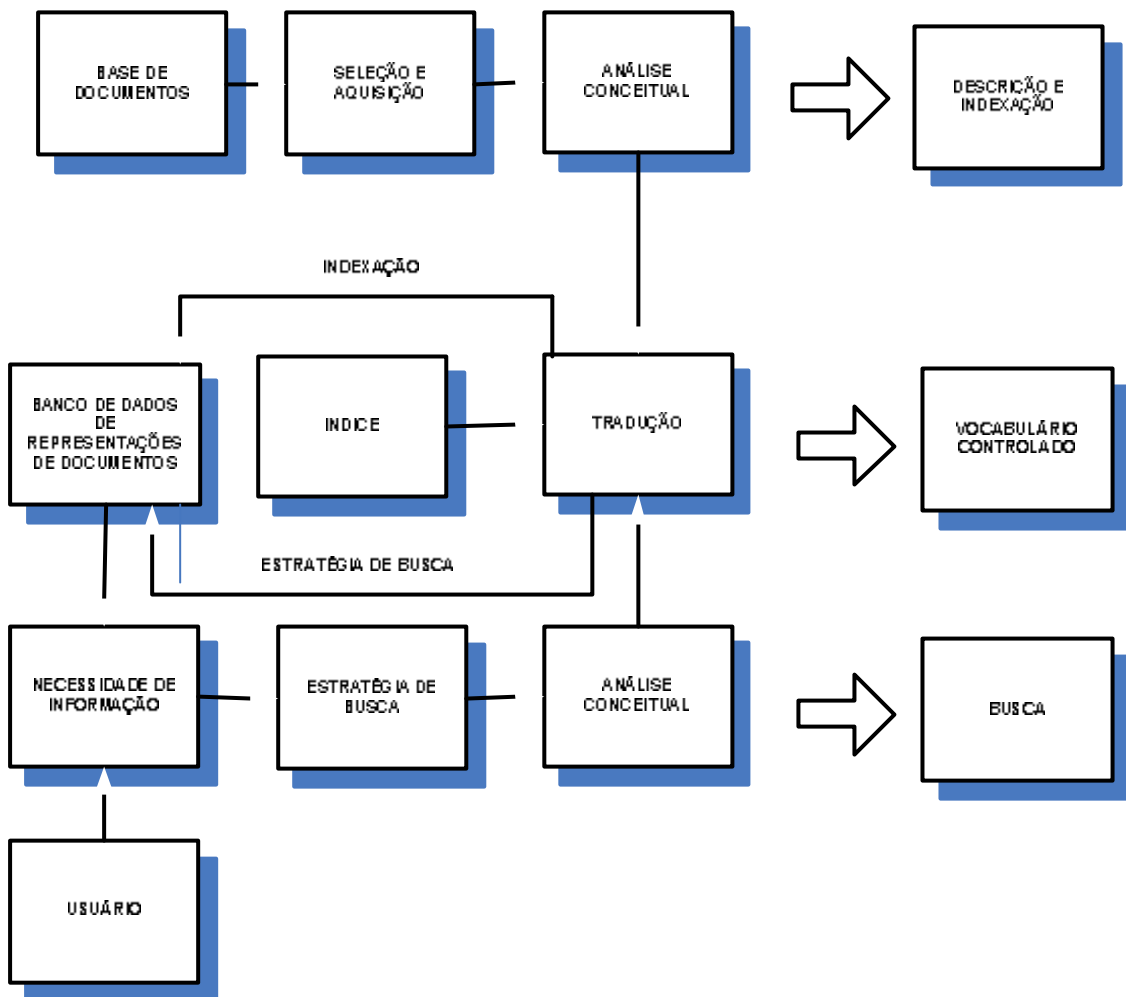
> *"esses sistemas oferecem maior número de pontos de acesso que os SRIs não automatizados, podendo-se, muitas vezes, pesquisar palavras-chave que aparecem em qualquer ponto do registro, inclusive no resumo e no texto completo, quando estes estão disponíveis. Além disso, permitem realizar pesquisas mais complexas, em que vários conceitos necessitam ser relacionados, pois pode-se combinar grande número de termos de busca com lógica booleana, de maneira que não seria possível nos impressos (p. 62)".*

Partindo de definições de vários autores, SOUZA (2006) afirma que os sistemas de recuperação da informação organizam e viabilizam o acesso aos itens de informação, desempenhando as atividades de:

> *"a) representação das informações contidas nos documentos, usualmente através dos processos de indexação e descrição dos documentos; b) armazenamento e gestão física e/ou lógica desses documentos e de suas representações; e c) recuperação das informações representadas e dos próprios documentos armazenados, de forma a satisfazer as necessidades de informação dos usuários. Para isso é necessário que haja uma interface na qual os usuários possam descrever suas necessidades e questões, e através da qual possam também examinar os documentos atinentes recuperados e/ou suas representações"*(p.163).

5.5.3 A Face da Gestão Documental

Em uma face da "moeda", conforme apresentado na figura 13, temos os documentos e suas necessidades de representação e armazenamento. Documentos são objetos (no contexto dessa tese, objetos textuais) potencialmente capazes de transmitir informações a prováveis usuários. É necessário que os documentos sejam selecionados, inseridos, descritos, representados, organizados e armazenados no SRI para que possam atender a potenciais necessidades dos usuários. Baseando-se em CENDON (2005) e LANCASTER (1979), podemos também apresentar um outro diagrama, adaptado abaixo, e que represente um sistema de recuperação da informação e que será analisado na seqüência, com foco na gestão documental.

BASE DE DOCUMENTOS
SELEÇÃO E AQUISIÇÃO
ANÁLISE CONCEITUAL
DESCRIÇÃO E INDEXAÇÃO
INDEXAÇÃO
BANCO DE DADOS DE REPRESENTAÇÕES DE DOCUMENTOS
INDICE
TRADUÇÃO
VOCABULÁRIO CONTROLADO
ESTRATÉGIA DE BUSCA
NECESSIDADE DE INFORMAÇÃO
ESTRATÉGIA DE BUSCA
ANÁLISE CONCEITUAL
BUSCA
USUÁRIO

futuros do sistema as quais julga digna de serem contempladas, seleciona e insere documentos na base de dados desse SRI, considerando critérios de autoria, assunto, tipo de documento, língua e origem. Em seguida, fará o processo representação do conteúdo temático dos documentos através da indexação, ou seja: é necessário determinar os pontos de acesso para a base de dados do sistema para que seja possível ao SRI responder às consultas dos usuários.

5.5.3.1 Representando os Documentos

Quanto maior e melhor a representação dos documentos, mais pontos de acesso temos. Esses pontos de acesso podem ser resumos ou termos. Em geral, os termos são mais utilizados, não obstante a importância dos resumos, embora, conforme o autor, listas de termos são como *miniresumos*. A representação, portanto, redunda na indexação, em geral indexação de assuntos, cujo processo envolve duas etapas: a análise conceitual e tradução (para um vocabulário ou linguagem de indexação). Para LANCASTER (2004), a análise conceitual implica em decidir o que trata o documento, ou seja: qual é o seu assunto. A tradução implica na conversão da análise conceitual de um documento num determinado conjunto de termos de indexação. Essa tradução pode levar ou não ao uso de um vocabulário controlado, como um tesauro ou lista de cabeçalho de assunto[41]. Um tesauro – vocabulário controlado – é o resultado da tradução dos conceitos e assuntos dos documentos da base em termos controlados. A indexação/tradução também pode ser feita a partir de termos livres, oriundos da linguagem natural, extraídos do próprio documento, tais como frases ou palavras que ocorrem em um documento ou escolhidos pelo gestor do SRI.

De acordo com ele, ao desenvolver a tradução, para indexar, temos dois tipos ações para a indexação, que pode ser por extração ou atribuição, conforme afirma:

> *"A esse respeito, faz-se uma distinção entre indexação por extração (indexação derivada) e indexação por atribuição. Na indexação por extração, palavras ou*

[41] De acordo com LANCASTER (2004), são três os principais tipos de vocabulários controlados: esquemas de classificação bibliográfica (como a CDD – Classificação Decimal de Dewey), as listas de cabeçalhos de assuntos e os tesauros. Todos procuram apresentar os termos tanto alfabética como sistematicamente. Os três tipos de vocabulário controlam sinônimos, distinguem homógrafos e agrupam termos afins, mas empregam métodos um tanto diferentes para alcançar tais objetivos. O autor faz um estudo mais amplo do assunto em LANCASTER, F.W. Vocabulary control for information retrieval. 2nd ed. Arlington, VA, Information Resources Press, 1986.

expressões que realmente ocorrem no documento são selecionadas para representar seu conteúdo temático. (...) A indexação por atribuição envolve a atribuição de termos ao documento a partir de uma fonte que não é o próprio documento. Os termos podem ser extraídos da cabeça do indexador; por exemplo, ele decidiria que os termos "AJUDA EXTERNA" e "RELAÇÕES EXTERIORES", que não aparecem explicitamente em nenhum dos resumos, seriam termos bons de usar no documento. (...) Mais frequentemente, a indexação por atribuição envolve o esforço de representar a substância da análise conceitual mediante o emprego de termos extraídos de alguma forma de vocabulário controlado"(LANCASTER, 2004. p. 18-19).

ROBREDO & CUNHA (1994) descrevem quatro critérios de análise relativos à atividade de indexação. Segundo esses autores, temos: a abrangência da análise conceitual, a estrutura a ser analisada no documento (partes dos documentos que serão analisadas), tipo de procedimento e o tipo de linguagem de indexação. No que tange ao primeiro critério, a abrangência, segundo esses autores, podemos realizar uma categorização, uma indexação superficial ou uma indexação profunda, sendo que:

- categorizar é trabalhar com a idéia de assuntos pré-existentes e em seguida reconhecer os aspectos dominantes dos documentos que fazem com que estes se enquadrem naqueles;
- indexar superficialmente é obter as principais idéias e conceitos a que um documento se refere;
- indexar profundamente é encampar todas as idéias e conceitos que são importantes.

Em relação ao segundo critério, a estrutura do documento, ROBREDO & CUNHA (1994) afirmam que a indexação pode se dar em relação no título, ao resumo, aos dois em conjunto, a partes determinadas tais como introdução ou conclusão e, por fim, em relação ao próprio documento completo. No terceiro critério, o procedimento de indexação pode ser manual, automático ou semi-automático (misto). Por fim, em relação à linguagem, a indexação pode ser em linguagem natural ou em linguagem artificial, sendo que:

- a linguagem natural pode ser livre (quando utiliza palavras extraídas do próprio documento), natural controlada não estruturada (quando usa listas de descritores padronizadas realizando eliminação de sinônimos e as listas de cabeçalhos de

assuntos), e natural estruturada (como um tesauro ou uma classificação facetada);

- a linguagem artificial pode ser artificial não estruturada (englobando esquemas de categorização que use qualquer símbolo não estruturado) ou artificial estruturada (englobando os esquemas de classificação bibliográficos, como a CDU – *Classificação Decimal Universal*).

Portanto, é por meio das características de um objeto de informação, ou dos documentos, que um SRI o pode tratar e representar para posteriormente o identificar e localizar esses mesmos documentos para atender as necessidades de informação do usuário. Assim sendo, é preciso uma visão lógica desses objetos de informação de forma a representá-los para que posteriormente possam ser dados como relevantes em uma interação com o usuário. Para BAEZA-YATES & RIBEIRO-NETO (1999) uma recuperação efetiva de informação relevante é diretamente afetada pelas duas tarefas do usuário, a busca e a navegação (analisadas na seção 5.5.4.2 e 5.5.4.3, respectivamente) e pela visão lógica dos documentos adotada no SRI.

5.5.3.2 Visão Lógica dos Documentos no Processo de Recuperação da Informação

Conforme BAEZA-YATES & RIBEIRO-NETO (1999) por motivos históricos, documentos em uma coleção não são armazenados diretamente num banco de dados, é preciso que estes passem por algum processo de abstração e são geralmente representados por um conjunto de termos de indexação ou palavras-chaves. Essas palavras podem ser automaticamente extraídas dos documentos ou selecionadas por especialistas humanos. De qualquer forma, isso gera uma visão lógica dos documentos. Computadores modernos estão tornando possível a representação de um documento por seu conjunto completo de palavras, é a chamada representação total do texto *(full text*). A representação *full text* é a mais completa visão lógica de um documento, mas seu uso implica em custos computacionais mais altos. Um conjunto menor de palavras selecionadas por especialistas humanos é a visão lógica mais concisa do documento, mas seu uso pode levar a uma recuperação de informação de baixa qualidade. Além disso, o SRI deve, também, reconhecer a estrutura normalmente presente em um documento, tais como: capítulos, seções, subseções, etc. A informação sobre a estrutura do documento deve ser usada nos processos de representação e recuperação em

diferentes modelos de sistemas de recuperação de documentos textuais. Esse processo, para documentos textuais, pode ser representado na figura 15, proposta por esses autores e reproduzida abaixo:

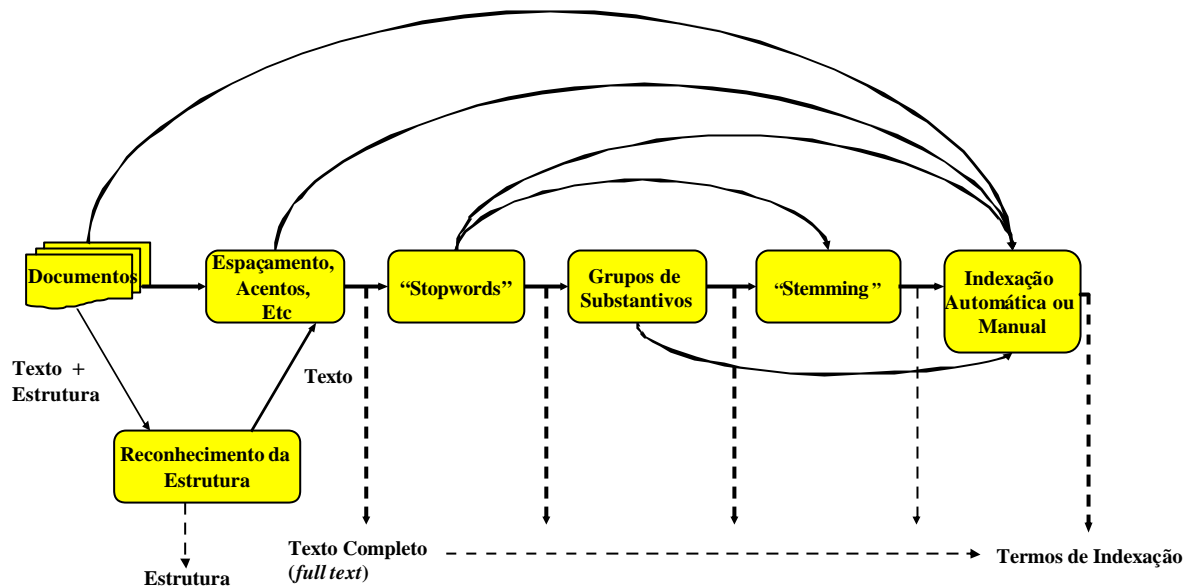



Figura 15: Visão Lógica do Documento: do texto completo ao conjunto dos termos de indexação. Fonte: BAEZA-YATES & RIBEIRO-NETO (1999) p.6

5.5.3.3 Processo de Recuperação da Informação

Existem muitas avaliações e representações do processo de recuperação da informação. BAEZA-YATES & RIBEIRO-NETO (1999), considerando documentos textuais, apresentam também uma visão sobre o processo de recuperação na figura 16 a seguir.

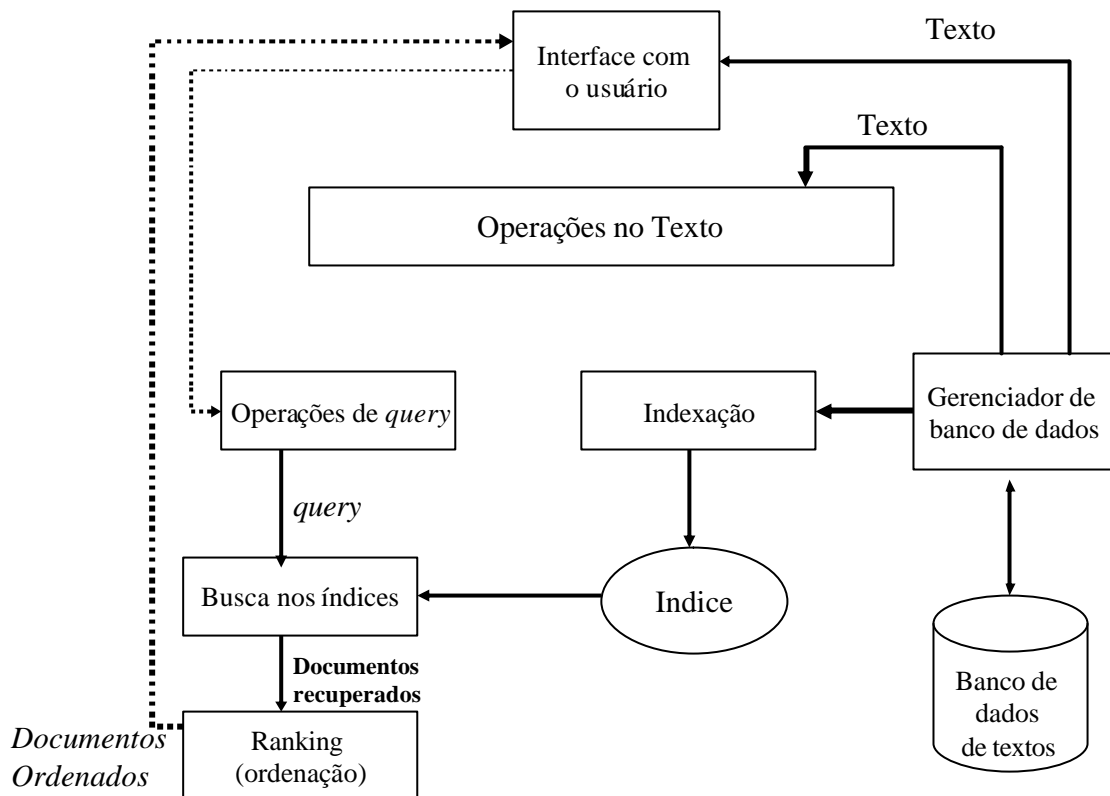



Figura 16: O Processo de Recuperação de Informações.
Fonte: BAEZA-YATES & RIBEIRO-NETO (1999) p.10

Nessa figura, os elementos representados significam o seguinte:

- *Interface com o usuário:* é a estrutura responsável para permitir a interação do usuário com os SRI's.
- *Operações no texto:* são operações realizadas sobre o texto de modo que seja reduzido o texto a ser pesquisado, a fim de que o processo de recuperação seja mais eficiente. Consistem em retirar artigos, conjunções, preposições, acentos e caracteres especiais e preposições que fazem elo entre as palavras, por exemplo: "linguagens de marcação" é o texto original digitado pelo usuário, porém, o texto real para a pesquisa ficaria "linguagens marcação".
- *Operações de Query:* são operações realizadas sob expressão ou palavra-chave utilizada pelo usuário para que o SRI possa fazer uma nova busca.
- *Busca:* é a estrutura que consiste em buscar na base de dados as informações contidas no módulo *Query operation.*

- *Ranking ou ordenação:* consiste em classificar os documentos recuperados na *Searching* por ordem decrescente de relevância e apresentá-los ao usuário.
- *Indexação:* é o processo de criação de índices para os documentos que serão adicionados à base de dados.
- *Índice:* são os documentos indexados.
- *Gerenciador da Base de Dados:* Gerente da base de dados de textos

5.5.3.4 Armazenando os Documentos

Concluído o processo de indexação/descrição, tais atividades criam segundo LANCASTER (2004), representações de documentos de um modo a poder haver sua inclusão em uma base de dados. Os próprios documentos se destinam a uma base de dados diferentes, ou seja: podemos ter a base de documentos e a banco de dados de representações dos documentos (informatizado ou não, como um recipiente de fichas representando os documentos). De acordo com CENDON (2005)

> *"A base de dados de documentos e a base de dados das representações de documentos hoje, já não são sempre distintas. Tornam-se cada vez mais comuns representações de documentos contendo não apenas termos de indexação, mas também resumos e, mais recentemente, tem se firmado a tendência das bases de dados conterem documentos em texto completo" (p. 64).*

Posteriormente, os documentos que forem relevantes para atender necessidades de informação para o usuário devem ser recuperados pelo SRI e disponibilizados para ele.

Os SRIs, realizam a atividade de armazenamento, tanto no nível lógico quanto físico dos itens de informação e em suas representações e os sistemas de arquivos são utilizados para tal armazenamento. Conforme CLAYBROOK (1987), um arquivo pode ser definido como um conjunto de registros que representam as entidades ou instâncias do mundo real através de seqüências de itens de dados (cada registro é composto de itens ou campos de dados). Tais itens são unidades de informação, representando os atributos e características das entidades do mundo real. Há distintos modelos para armazenamento de arquivos. SALTON & MCGILL (1983) descrevem, por exemplo, os

modelos de arquivos seqüenciais, arquivos seqüenciais ordenados e arquivos indexados (os quais se subdividem em arquivos diretos e arquivos invertidos).

5.5.4 A Face da Gestão das Relações com o Usuário

5.5.4.1. A Questão da Relevância e da Atinência no Atendimento ao Usuário

A noção de relevância é o centro da recuperação da informação. Realmente, o objetivo principal de um SRI é, da melhor forma possível, recuperar todos os documentos que são relevantes para uma consulta do usuário, ao mesmo tempo em que recupera poucos documentos não relevantes.

Para LANCASTER (2004) são sinônimas as expressões "útil", pertinente e relevante para uma necessidade de informação. Ou seja: um documento pertinente (útil) é aquele que contribui para satisfazer uma necessidade de informação. Esse documento pode contribuir na medida em que é atinente, ou seja: trata de assunto ou assuntos que respondem às necessidades de informação do usuário. Para ele:

> *"O tema da atinência está relacionado muito de perto com o da relevância – isto é, a relação entre um documento e uma necessidade de informação ou entre um documento e um enunciado de necessidade de informação (uma consulta)"* (p. 14).

MIZZARO (1997) realizou um estudo amplo sobre a questão da relevância, tendo identificado a ocorrência de uma vasta literatura sobre este tema em diversas áreas (filosofia, psicologia, lingüística, entre outras). Para esse autor, a informação relevante está diretamente relacionada com o usuário, com a necessidade de informação que ele possui bem como o momento que isso ocorre. Segundo ele, a relevância pode ser vista como o relacionamento entre duas entidades, uma de cada um dos seguintes grupos:

- Grupo constituído do que denomina de entidades *documento* (o objeto que o usuário vai obter depois de sua busca), sua *representação* ou *surrogate* (as palavras-chave, por exemplo) e *informação* (o que o usuário recebe quando lê um documento);

- Grupo do que classifica como entidades *problema* (que necessita de informação para ser resolvido*), necessidade de informação* (o que o usuário entende ou percebe do problema – representação do problema na mente do usuário*), solicitação* ou *requisição* (representação da necessidade de informação do usuário em uma *linguagem humana* – geralmente *linguagem natura*l) e *consulta* ou *query* (a representação da necessidade de informação do usuário na *linguagem do sistem*a).

Em outro trabalho, MIZZARO (1998) propõe uma estrutura em quatro dimensões que detalham o conceito de relevância no processo de recuperação de informação. Essas dimensões são: os recursos de informação, a representação do problema do usuário, o tempo e os componentes.

Os componentes da primeira dimensão, os "recursos de informação" são: o documento (entidade física que uma pessoa vai obter depois da busca da informação) a representação (feita por titulo, palavras chaves, autor, resumo ou dados bibliográficos) e a informação (entidade não física que o usuário produz ao ler/examinar um documento).

A representação do problema do usuário é a segunda dimensão. Ela é caracterizada por quatro pontos: necessidades reais e necessidades percebidas do usuário, solicitação, que é a representação de uma necessidade de informação em linguagem humana e, na seqüência temos a consulta, que é a expressão da necessidade percebida de informação na linguagem de um SRI.

O tempo é a terceira dimensão. Um documento pode ser relevante em relação ao momento em que o usuário o examina. Ele pode não ser relevante agora e ser depois, ou o contrário.

Os componentes configuram a quarta dimensão, sendo eles:

- o aspecto *tópic*o, que se refere ao assunto que o usuário deseja;
- a *tarefa*, que se refere ao que o usuário vai fazer com os documentos retornados; e

- o *contexto*, que inclui tudo aquilo que não pertence aos tópicos anteriores, mas, de alguma forma, afeta a busca e a avaliação dos resultados, tais como documentos recuperados já conhecidos, tempo ou dinheiro disponível para a busca, situação na qual a busca é realizada, compreensibilidade da informação recuperada, dentre outros.

Para MONEY (2006) a relevância é resultado de um julgamento subjetivo, podendo incluir:

- a realização das expectativas do usuário, no sentido de satisfazer os usos desejados, atendendo, portanto, suas necessidades de informação;
- a pertinência no assunto que se deseja;
- a atualidade (ser informação recente);
- a confiabilidade (de uma fonte que mereça crédito).

Conforme LANCASTER (2004), há seis critérios para avaliação de SRIs:

- Precisão
- Revocação
- Tempo de Resposta
- Esforço do usuário para obter resultados que atendam suas necessidades
- Formato de apresentação dos resultados
- Exaustividade (quanto uma coleção em questão contém de documentos potencialmente relevantes)

Os dois primeiros critérios, precisão e revocação, medem a eficácia do sistema na recuperação de informações relevantes. A precisão é a taxa de documentos relevantes recuperados sobre o total de documentos recuperados, isto é: a proporção de material recuperado que é efetivamente interessante. Por outro lado, a cobertura ou revocação é uma taxa ou coeficiente de documentos relevantes recuperados sobre o total de documentos relevantes existente na base, ou seja: a proporção de material relevante recuperado em relação ao que existe no acervo ou base. As idéias de precisão e revocação podem ser representadas na figura 17 abaixo:

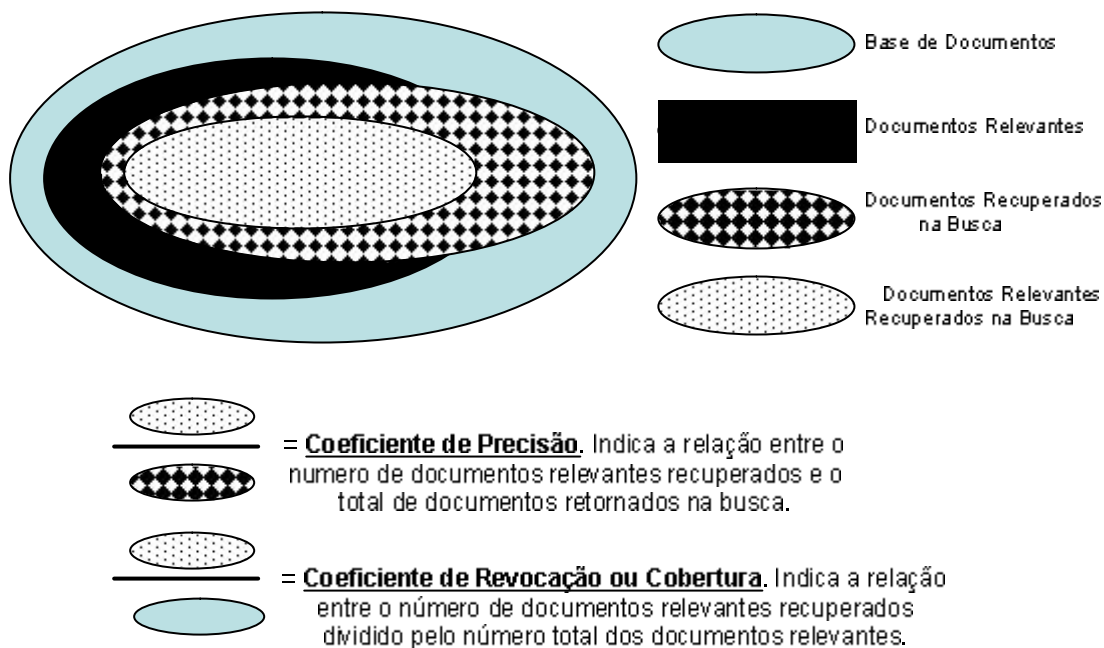



Figura 17: O problema da recuperação da informação e as idéias de precisão e revocação. Baseado em LANCASTER, 2004, p3

O papel dos SRIs é ajudar os usuários a descobrir quais documentos são relevantes para ele em uma coleção, servindo-se para isso de visões e modelos diferentes para atender a tal necessidade. De acordo com BAEZA-YATES & RIBEIRO-NETO (1999) nos sistemas automatizados isso é realizado por algum algoritmo que, atuando com base em alguma heurística previamente definida pode "escolher" os documentos que considera relevantes para recuperação e os ordena de acordo com critérios previamente estabelecidos no desenho ou proposta conceitual do sistema.

5.5.4.2 A Necessidade de Informação

De acordo com a perspectiva de BURNETT & MCKINLEY (1998) sobre os *"Estados Anômalos de Conhecimento" (Anomalous State of Knowledge*), que pode ser representativa de grande parte dos modelos de SRIs existentes em suas suposições, representações e natureza, uma pessoa utiliza um sistema de informação a fim de sanar uma anomalia pessoal de conhecimento. A anomalia é tratada quando a informação recuperada satisfaz o estado de anomalia individual de conhecimento.

Realmente, usuários se relacionam com os SRIs a partir de suas necessidades de informação as quais buscam responder. Responder a uma necessidade de informação é encontrar a informação relevante em base de documentos a partir de estratégias de busca. Uma informação relevante. Relevante, portanto, é tudo o que pode satisfazer a certas necessidades de informação dos usuários. As relações com o usuário passam pelas questões da tarefas do usuário e de sua interface com o SRI.

O SRI é um conjunto de atividades integradas capazes de promover o casamento entre as necessidades do usuário com as informações recuperadas na base de documentos, gerando, portanto, resultados relevantes para busca desse usuário. Um usuário é uma pessoa que possui alguma necessidade de informação e adota uma postura de busca dessa informação para satisfazer tal necessidade. Uma necessidade de informação, de acordo com MIZZARO (1996), é uma carência de conhecimento que um usuário tem para realizar uma determinada ação. SARACEVIC (1983) citado por ARAÚJO (2004) pontua cinco elementos básicos em situações de busca de informações pelos usuários, quais sejam:

- o problema
- o estado interno de conhecimento
- a intenção
- a questão
- o pedido.

SHNEIDERMAN (1997) apresenta quatro fases para busca de informação:

- Formulação: expressão da busca;
- Ação: lançamento da busca;
- Avaliação dos Resultados: leitura das mensagens e saídas;
- Refinamento: formulação do próximo passo.

REES & SARACEVIC (1963) lembram que inicialmente tem-se um problema de informação que é traduzido em necessidade de informação. Essa necessidade gera uma pergunta, expressando um conceito mais concreto. Pode acontecer que a pergunta

não expresse a necessidade de informação. A pergunta redunda em uma solicitação ao SRI, sendo que, novamente, nem sempre a pergunta que o usuário possui na mente é a que é expressa explicitamente na sua solicitação. As perguntas embutem aspectos de significado e linguagem. A partir da pergunta, temos a formulação de uma estratégia de busca para atender à necessidade do usuário.

BAEZA-YATES & RIBEIRO-NETO (1999) abordam a questão da interface do usuário com o sistema e o processo de acesso à informação. Avaliam a importância de um adequado processo de interação (*interface*) entre o usuário e o SRI para que o processo de recuperação da informação funcione adequadamente. De acordo com esses autores, o processo de acesso à informação se inicia com a manifestação de uma necessidade de informação. A partir daí, o usuário seleciona um sistema e as coleções para realizar buscas sobre elas. Para isso, ele formula uma consulta e a submete no SRI. Recebe então os resultados em forma de itens de informação, avaliando e interpretando os resultados. Ele pode encerrar o processo ou, se os resultados não forem satisfatórios, recomeçar tudo outra vez.

Nos SRIs, o projeto de interface é essencial para o uso eficiente do sistema. A interface deve atender a diferentes tipos de usuários com a gama completa de tarefas que o usuário poderá procurar para executar com esse sistema. O processo de interface tornou-se mais importante com o aumento e variedade de usuários. A interação é feita através dos dispositivos de entrada e saída do computador (SOUZA, 2006).

Retomando a perspectiva de LANCASTER (2004) e partindo da premissa da existência de bases de documentos, impressos ou eletrônicos, o ciclo da recuperação da informação se fecha com o atendimento às necessidades de informação dos usuários finais ou comunidade de usuários. A partir de suas necessidades, os usuários definem suas estratégias de busca, que podem ser simples, tal como a escolha de um único termo para consultar uma determinada base ou envolver muitas combinações em estratégias mais complexas. Essa busca significa, em geral, na localização dos conceitos, dos temas, dos assuntos requeridos pelo usuário, cujos termos, em geral, representam sua explicitação. Se o SRI em questão utilizar um vocabulário controlado, as consultas dos usuários devem ser analisadas conceitualmente e traduzidas para o vocabulário do sistema.

As estratégias de busca podem ser consideradas, de acordo com BAEZA-YATES & RIBEIRO-NETO (1999), dentro de tarefas ou atividades realizadas pelo usuário do SRI. Os autores em questão postulam que o usuário possui dois papéis no processo de recuperação da informação: recuperação ou busca e navegação, ou "*brauseio*", conforme ARAÚJO (1994). A busca ocorre em geral quando o usuário possui objetivos mais bem definidos, enquanto a navegação é uma estratégia em que os objetivos do usuário não estão bem definidos. SRIs clássicos normalmente permitem recuperação de informação rápida. Por outro lado, sistemas de hipertexto são geralmente criados para permitir navegação rápida. Bibliotecas digitais modernas e interfaces para a *web* devem tentar combinar essas duas tarefas, sendo essa uma tendência.

SHNEIDERMAN (1997), por sua vez, relaciona tipos de tarefas de busca de informação:

- Busca de fatos específicos (busca de item conhecido): por exemplo, encontrar a informação de um livro específico de um determinado autor;
- Busca de fatos estendida: por exemplo, que outros livros do autor de uma determinada publicação existem?
- *Browsing*: por exemplo, há novo estudo sobre os peixes marinhos em Angra dos Reis?
- Exploração de disponibilidade: por exemplo, que informação existe sobre a mata atlântica no sul do Rio de Janeiro?

Trataremos os dois primeiros itens como tarefa de busca e os dois últimos como tarefa de navegação, a seguir

5.5.4.3 A Tarefa da Busca

Partindo de sua necessidade, o usuário de um SRI tem que traduzi-la em uma consulta, escrita na linguagem fornecida pelo sistema. Geralmente, isso implica em especificar um conjunto de palavras que conduzam à semântica de sua necessidade.

Nesse caso, o usuário está buscando por informações úteis executando uma tarefa de recuperação. Existe uma distinção entre duas tarefas que podem ser executadas pelo usuário de um SRI: a recuperação de informação ou a navegação entre documentos, como mostra a Figura 18

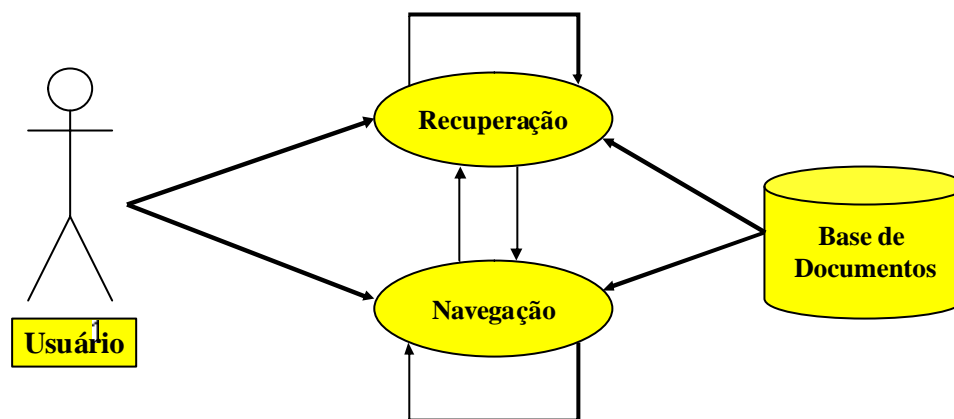



Figura 18: As Tarefas do Usuário no uso de um Sistema de Recuperação da Informação
Fonte: BAEZA-YATES & RIBEIRO-NETO (1999), p. 4

Conforme CENDON (2005):

> *"é elaborada a estratégia de busca e formulada a expressão de busca, na qual os termos da busca são relacionados entre si através de operadores booleanos ou não booleanos. A estratégia de busca consiste em um plano para encontrar a informação desejada em que várias expressões de busca podem ser utilizadas. Através da expressão de busca, o sistema compara, então, as representações dos documentos com as das perguntas dos usuários. Na fase final, os documentos recuperados através da consulta ao sistema são apresentados ao usuário para que julgue, então, sua relevância para as suas necessidades de informação.".*

SOUZA (2006), baseando-se em BAEZA-YATES & RIBEIRO-NETO (1999), lembra que a recuperação pode se dar "*ad hoc*" e por filtragem. Conforme suas palavras:

> *"Quando o acervo de documentos sofre poucas alterações enquanto novas queries são submetidas ao sistema, chama-se o modo de operação de "recuperação ad hoc". Quando as queries se mantêm relativamente estáticas enquanto novos documentos são adicionados, chama-se a esse modo de operação de filtragem (filtering). A filtragem acontece usualmente em processos de monitoração de fontes de informação, enquanto a recuperação ad hoc representa as buscas usuais em SRIs"* (p. 165-166).

5.5.4.4 A Tarefa da Navegação

A navegação é "a arte de não saber o que se quer até se encontre", conforme ARAÚJO (1994) que lembra também que essa tarefa é fortemente visual e tem um forte componente de acesso direto, podendo ser associado com formas e padrões em termos de imagens e distribuição do texto numa página ou numa tela de computador. Essa autora cita COVE & WALSH (1988), para os quais a navegação é uma ação do usuário de caminhar não com um objetivo claramente definido, mas com uma noção ou referenciais vagos de interesse. Ele anda a esmo sobre uma coleção de documentos, recolhendo ou examinando alguns deles, ora superficialmente, ora com um pouco mais de detalhe. Também apresenta a contribuição de MORSE (1973), que afirma ser a navegação um tipo de busca semelhante à busca de um navio, realizada por um observador em um avião. Esse observador não sabe ao certo se aquilo que está a procurar, o alvo, está lá ou não, e mesmo de que ele perceberá esse alvo se lá ele estiver, mas o vôo ajuda muito.

Para BAEZA-YATES & RIBEIRO-NETO (1999), na navegação o usuário não propõe uma questão ou realiza uma busca *(quer*y), que represente uma necessidade de informação ao sistema. mas, navega através dos documentos – que não foram necessariamente indexados previamente – buscando informações de interesse. Ainda de acordo com os autores, temos três tipos de navegação: a plana (*flat browsing*), a guiada pela estrutura (*structure guide browsing*) e a do modelo de hipertexto (*hipertext model*). Uma navegação hipertextual é, por exemplo, aquela realizada em estruturas de arquivos ligados em rede.

O aumento das coleções digitais e a grande difusão do hipertexto, sobretudo com a expansão da Internet e das redes corporativas, demandam e ampliam hoje as necessidades e as possibilidades da tarefa de navegação, sobretudo com o advento dos softwares de navegação, tais como *Netscape*, *Explorer*, *Mozzila*, além dos recursos inseridos em ferramentas de colaboração e de portais corporativos de organizações diversas, tais como *Lótus Notes* e *Verity*. Cumpre salientar também que a navegação não se resume às coleções digitais, mas pode ser também realizada em acervos físicos, como em uma biblioteca.

Em um sentido mais estrito, a navegação é a forma como os usuários se movimentam através de um hiperdocumento. Ela pode ser influenciada pela estrutura do documento, pela estratégia de navegação e pelos mecanismos auxiliares de navegação que o usuário possa encontrar no SRI.

5.6 Modelos de Recuperação da Informação

É possível propor diferentes classificações para os SRIs, com base em diferentes critérios e objetivos, da mesma forma que há diferentes possibilidades de se classificar um sistema de informação, conforme já discutido anteriormente. Essas classificações dão origem a modelos que são classificações conceituais ou abordagens genéricas visando recuperar informações. Conforme já dito na presente tese, de acordo com LANCASTER (1968), de acordo com o tipo de aplicação que se dá aos Sistemas de Recuperação da Informação – SRIs, eles podem ser classificados em:

- sistemas de recuperação de documentos;
- sistemas de recuperação de referencias;
- sistemas de recuperação de dados e:
- sistemas de recuperação de fatos (ou "*query answering systems*").

WIVES (2002) propõe uma classificação derivada desse modelo em que os SRIs poderiam ser classificados em:

- sistemas de recuperação bibliográfica;
- sistemas de recuperação de informação textual;
- sistemas de recuperação de informação visual;
- sistemas para bibliotecas digitais.

No detalhamento dos modeles de recuperação textual esse autor enumera como modelos de recuperação:

- *booleano*;
- espaço vetorial;

- probabilístico;
- difuso;
- busca direta;
- aglomerados;
- lógico;
- contextual.

Sob a perspectiva da estrutura do documento, em geral todos os documentos textuais tendem a ser considerados documentos não estruturados em oposição aos dados estruturados de sistemas relacionais, sobretudo por serem escritos em linguagem natural livre. Entretanto, com um olhar mais atento, é necessário fazer algumas ressalvas. Muitos documentos textuais possuem sim elementos estruturais significativos que podem permitir a representação e a recuperação mais eficientes. Então, não em relação à outro tipos de objetos informacionais (pois não se pode comparar a intensidade de estruturação de um texto com uma tabela de banco dados relacional), mas em relação aos próprios textos, podemos classificar os documentos em textos estruturados e textos não estruturados. Os documentos de textos estruturados são aqueles em que é possível reconhecer elementos estruturais e de semântica bem definida, enquanto os não estruturados correspondem a textos livres, com pouco ou nenhum formato, bem como estrutura facilmente identificável. Por isso uma forma de classificar os modelos de recuperação da informação é se eles visam a documentos de texto estruturados ou de textos não estruturados. Sendo assim, teríamos a seguinte taxonomia:

- Modelos de recuperação para textos não estruturados: booleano, vetorial e probabilístico.

- Modelos de recuperação para textos estruturados: listas não-sobrepostas, proximidade de nós e recuperação por passagens.

Uma outra taxonomia que ilustra quinze modelos de Recuperação da Informação é apresentada por BAEZA-YATES & RIBEIRO-NETO (1999), conforme a Figura 19 a seguir.

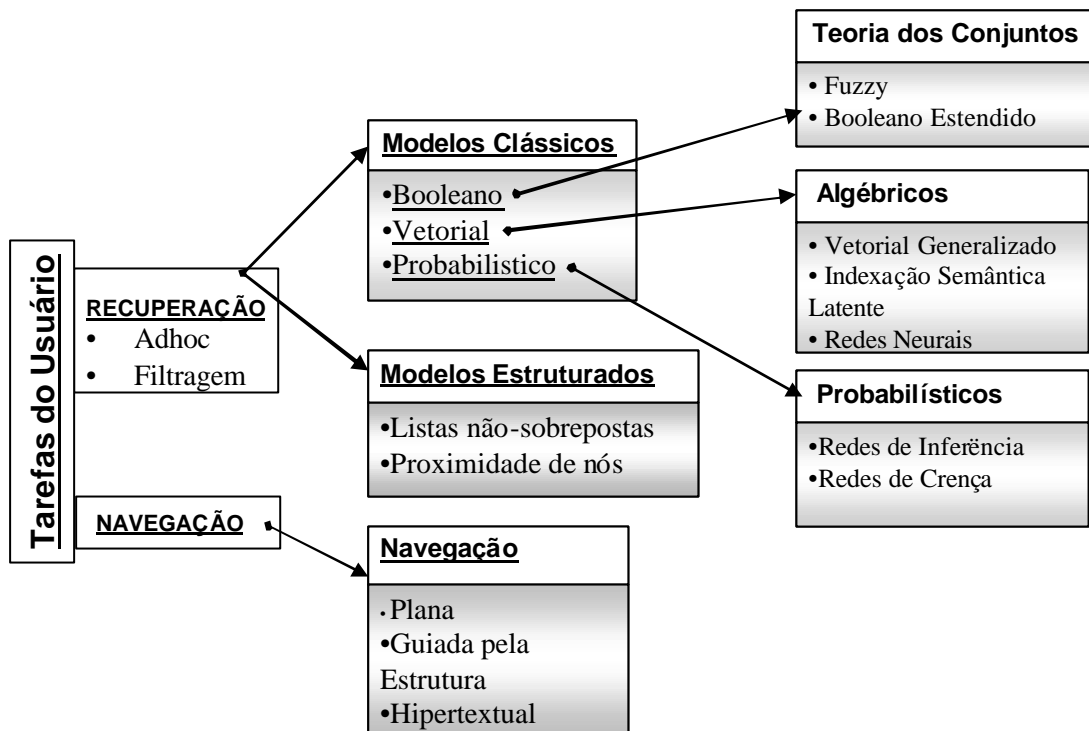



Figura 19: Uma Taxonomia dos Modelos de Recuperação da Informação
Fonte: BAEZA-YATES & RIBEIRO-NETO (1999), p. 21

Uma breve explanação sobre tais modelos, a partir da análise da concepção desses autores, e algumas outras contribuições pode ser sintetizada a seguir.

5.6.1 Modelos Clássicos

Os modelos clássicos de recuperação possuem estratégias de busca de documentos relevantes para uma consulta *(query*). Eles são três: *booleano,* vetorial e probabilístico e para cada um deles há desdobramentos possíveis que visam melhorá-los em seus propósitos e na sua performance. Tanto a consulta feita pelo usuário, quanto os documentos que compõem a coleção a ser pesquisada, são representados pelos seus termos. Portanto, tal abordagem assume que cada documento é descrito por um conjunto de palavras chaves, chamadas termos de indexação. Nem todos os termos de um documento são úteis para representar seu conteúdo. Os termos com freqüência menor permitem identificar conjuntos mais restritos. A *importância* de um índice é representada pelos pesos a ele associados. Se considerarmos que *ki* é um índice e *dj* é um documento, *wij* é o peso associado a *(ki,dj),* que quantifica a importância do índice na descrição do conteúdo do documento, conforme especificado por BAEZA-YATES & RIBEIRO-NETO (1999).

5.6.1.1 Modelo *Booleano*

É um dos modelos clássicos que define uma consulta como uma expressão *booleana* convencional, que liga seus termos através de conectivos lógicos de Boole (*AND, OR e NOT*). Considera os documentos como conjuntos de palavras e manipula e descreve esses conjuntos com base em tais conectivos. No modelo *booleano* um documento é considerado relevante ou não relevante para uma consulta, não existindo resultado parcial e não havendo informação que permita a ordenação do resultado da consulta, uma vez que possui a desvantagem de trabalhar com a lógica binária. É um modelo muito mais utilizado para recuperação de dados do que para recuperação de informação. A passagem da necessidade de informação do usuário à expressão *booleana* é considerada complicada: o usuário precisa conhecer álgebra *booleana*, o que na maioria dos casos, não acontece. Além disso, as consultas *booleanas* formuladas pelos usuários são freqüentemente simplistas, o que faz com que exista um retorno de poucos ou de muitos documentos em uma consulta. Apesar das restrições, o modelo é bastante utilizado em sistemas comerciais.

É possível melhorar os resultados gerados nesse modelo, usando conjuntos denominados difusos ou nebulosos (*fuzzy*), atribuindo pesos aos termos, dividindo a consulta em classes e conceitos, trabalhando com proximidades dos termos entre si num documento e identificando mais precisamente que termos são melhores para definir um consulta. Ele possui alguns desdobramentos conforme se segue.

a) **Modelo Nebuloso (*Fuzzy*)**

Esse modelo trabalha com a teoria dos conjuntos em que a pertinência ou não de um elemento a ele pode variar entre 0 e 1. Essa teoria de conjuntos difusos (fuzzy), é baseada em ZADEH (1965) para quem as características de determinado universo estão presentes em todos os conjuntos, mas tal presença , além de poder ser medida, e pode não ser exata, ou seja, pode haver incerteza. Assim sendo, não temos, por exemplo, um conjunto vazio, mas sim um conjunto cujos elementos possuem uma relevância muito baixa (próxima de zero). Na visão desse autor, é muito severa a idéia de que um objeto pertence ou não a um conjunto, ou seja: não há possibilidade do objeto pertencer

parcialmente ao conjunto. Na prática, as pessoas utilizam raciocínios onde um objeto pode pertencer parcialmente ao conjunto. Um exemplo é quando usamos expressões do tipo “mais ou menos”, “muito”, “pouco” e “talvez”. Considerando que tem como base uma visão teórica que manipula conjuntos, esse modelo é considerado uma extensão do modelo *booleano*. É possível utilizar um como um tesauro para determinar que termos relacionados semanticamente aos termos índice também possam determinar algum grau de pertencimento ao conjunto difuso determinado pela *query*.

b) Modelo *Booleano* Extendido

É o modelo que é desenhado com atribuição de pesos aos termos, uma vez que leva em conta a importância das palavras nos documentos. A carga semântica dos termos é completamente diferente, quando se tem, por exemplo, uma consulta com dois termos ou duas consultas distintas com um termo cada. O usuário pode especificar a importância de um termo na consulta, embora isso possa aumentar para ele o grau de complexidade para uso desse tipo de sistema.

5.6.1.2 Modelo Vetorial

Esse modelo foi desenvolvido e descrito por SALTON & MCGILL (1983) para ser utilizado num SRI chamado SMART. No modelo vetorial, cada documento é representado como um vetor de termos. Cada termo possui um valor associado que indica o grau de importância (peso) deste termo em um documento. De acordo com WIVES (2002), as distâncias entre um documento e outro indicam seu grau de similaridade. Por isso, o vetor resultante para uma consulta específica é definido através de um cálculo de similaridade. São atribuídos pesos aos termos das consultas e documentos que determinam o tamanho e a direção de seu vetor de representação. O ângulo formado por estes vetores recebe o nome de q. O *cos* de q determina a proximidade da ocorrência. O cálculo da similaridade é baseado neste ângulo entre os vetores que representam o documento e a consulta (Figura 20). Isso implica que os documentos que possuem os mesmos termos acabam sendo colocados em uma mesma região do espaço e, em teoria, tratam de um assunto similar (característica que confere o nome espaço-vetorial ao modelo). As consultas dos usuários podem também ser expressas por um vetor, o que faz então com que possam ser comparados os vetores de

documentos e os de consulta. O grau de similaridade entre cada um deles pode ser identificado. Os documentos mais similares, próximos ao espaço da consulta são mais relevantes para os usuários e devem ser retornados como resposta a ela.

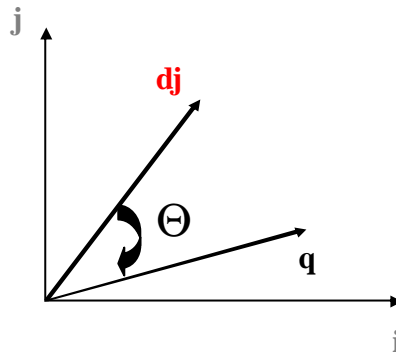



Figura 20: O Espaço Vetorial

Conforme SOUZA (2006):

> *"nesse modelo, os documentos são modelados como "sacos de palavras" (bags of words), e são representados como vetores no espaço n-dimensional, onde n é o total de termos índices (palavras) de todos os documentos no sistema. No modelo, que é não binário, pode-se calcular um grau de similaridade a ser satisfeito pelos documentos para serem considerados relevantes (ex: que as palavras apareçam ao menos duas vezes, etc.) e determinar o grau de similaridade, com vistas a construir um ranking. O modelo vetorial é a base da grande maioria de sistemas de recuperação de informações, mais notadamente os que têm como objeto a Internet, embora estes utilizem também outras técnicas para determinar o ranking de documentos como resposta a uma consulta".*

Dentre as vantagens do modelo vetorial podemos dizer que:

- atribuir pesos aos termos leva a uma melhora o desempenho;
- é um processo de encontro parcial (função de similaridade), sendo melhor do que a exatidão do modelo *booleano*;
- documentos são ordenados de acordo com seu grau de similaridade com a consulta.

No que tange às suas desvantagens podemos pontuar:

- a ausência de "ortogonalidade" entre os termos pode fazer encontrar relações entre termos que aparentemente não têm nada em comum;
- ele é um modelo generalizado;
- um documento relevante pode não conter termos da consulta.

Alguns outros modelos possuem a proposta de desdobrar o modelo vetorial acima descrito, tais como o vetorial generalizado, a indexação semântica latente e as redes neurais. São eles o vetorial generalizado, a indexação semântica latente e as redes neurais.

a) Vetorial Generalizado

Também conforme SOUZA (2006):

> *"nesse modelo, questiona-se a independência dos termos índices, assumida nos modelos clássicos, e abre-se a possibilidade de considerar que certos conceitos – representados por estes termos – sejam relacionados.Uma das formas de determinar relações entre termos é examinar a co-ocorrência desses no texto de cada documento, além do exame das relações semânticas estabelecidas por um tesauro, como foi comentado" (p. 167).*

b) **Indexação Semântica Latente**

O grande aumento do volume de textos digitais e a sofisticação das necessidades de informação dos usuários provocaram um maior interesse em processos de busca que levem em conta o conteúdo dos objetos de informação (os documentos), valorizando seu entendimento semântico para aprimorar as respostas às consultas formuladas pelos usuários. Respondendo a esse desafio, a indexação semântica latente promove a identificação de conceitos nos textos de modo independente dos termos que os representam. Ela viabiliza encontrar uma estrutura semântica associada a uma coleção de documentos. Mesmo que existam palavras diferentes expressando uma mesma idéia, esse método consegue detectar as idéias ou conceitos-chave presentes e identificá-los. De acordo com ANDERSON & PÉREZ-CARBALLO (2001) citado por LANCASTER (2004), a indexação semântica latente (ISL):

"é um dos mais elaborados esforços atuais visando a uma indexação automática de qualidade. Fundamenta-se em agrupamentos de termos baseados em co-ocorrência e identificação de documentos relativos a tais agrupamentos. Ao se apoiar em dados de co-ocorência a ISL também consegue lidar com o problema da variedade de termos que expressam idéias semelhantes. Como exemplo da capacidade da ISL lidar com documentos diferentes, imaginemos documentos sobre conserto e manutenção de automóveis. Documentos diferentes usarão vários termos diferentes como "automóvel", "carro", "veículo", "sedã", além dos nomes das marcas e modelos (...). O programa (baseado em ISL) relacionará esses termos entre si devido ao alto nível de co-ocorrência com termos como óleo, gasolina, combustível, carburador, pneus, ar-condicionado, etc. Ele cria agrupamentos de termos altamente relacionados (por meio da co-ocorrência), de modo que, quando um número suficiente deles ocorre num documento, este pode ser ligado ao agrupamento respectivo. Assim, é possível fazer buscas sobre cuidado e manutenção de carburadores de automóveis a gasolina sem nos preocuparmos com as palavras específicas usadas para automóvel. Todas as palavras que significam mais ou menos o mesmo que automóvel serão ligadas ao mesmo agrupamento". (p. 266).

Embora possa ser visto como um método para realizar indexação automática, a ISL, de acordo com LANCASTER (2004) é muito mais uma maneira de desenvolver automaticamente uma estratégia de busca para produzir termos semanticamente relacionados. Por exemplo, de acordo com ele:

"o termo A estará um tanto relacionado com o termo Y se ambos ocorrerem frequentemente com o termo Q. Com esse método, poder-se-á recuperar documentos possivelmente relevantes cujos termos de indexação diferem dos termos da consulta mas estão estatisticamente relacionados a ele". (p. 316).

GORDON & DUMAIS (1998) afirmam que:

"*Na prática isso significa que dois documentos que usam vocabulários com alto grau de duplicidade podem ser ambos recuperados mesmo que a consulta somente empregue os termos que indexam um deles. Igualmente, termos serão considerados ´próximos` uns dos outros se ocorrerem em conjuntos de documentos coincidentes*" (p. 677).

A Indexação Semântica Latente (ISL) procura trabalhar com um espaço menor do que os espaços *booleano* ou vetorial de palavras. Esse espaço dimensional menor é porém mais significativo, implicando em uma análise sinóptica das correlações de

termos em uma coleção de documentos. Podemos estabelecer o seguinte esquema para a ISL, expresso na Figura 21:

- Um termo (T1, T2 ou T3), representando um conceito ou idéia de interesse, ativa um espaço conceitual.
- O espaço conceitual caracteriza uma perspectiva, uma visão, um contexto, uma idéia central em torno do qual outras idéias ou eventos se articulam e se fazem presentes.
- Esse espaço conceitual é expresso por palavras, termos ou trechos em documentos em uma base de documentos de interesse (D1, D2, D3, D4...).

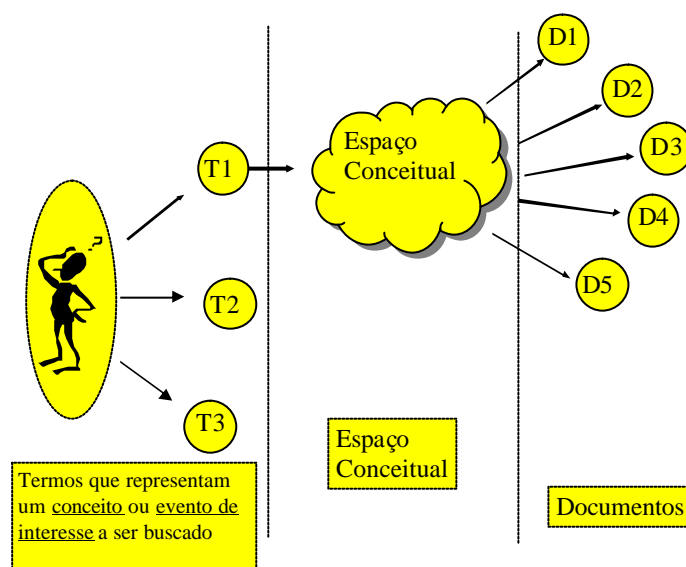



Figura 21: Uma representação da Indexação Semântica Latente
Fonte: Ilustração elaborada pelo autor

A ISL procura responder à necessidade de análise semântica no tratamento da informação. Auxilia a encontrar uma relação semântica entre uma consulta e os termos indexados. A ISL consegue recuperar documentos mesmo nas situações em que não se encontrem palavras idênticas às da consulta. De acordo com FORONDA (2005):

> *"A Indexação Semântica Latente (do inglês Latent Semantic Indexing, LSI) tenta superar as deficiências da recuperação por combinação de termos, tratando a falta de confiabilidade dos dados associados a uma relação termo-documento ou documento-documento, como um problema estatístico. Este método assume que há uma estrutura semântica oculta (latente), subjacente aos dados. Esta semântica é esquecida parcialmente pela aleatoriedade da escolha da palavra no que se refere à recuperação, pelo fato de que se escolhem*

palavras individuais para serem recuperadas, indexadas, etc. Utiliza-se no LSI um modelo matemático para estimar esta estrutura latente, que liberta do "ruído" constituído pela polissemia e pela sinonímia existente nos documentos. A descrição dos termos e dos documentos baseados na estrutura semântica latente é utilizada tanto para a indexação como para a recuperação. Entende-se por "estrutura semântica" a estrutura de correlação entre as palavras individuais que aparecem nos documentos; "semântico" implica o fato de que os termos, em um documento, possam ser tomados como referentes ao documento ou ao assunto desse documento. Esta técnica de análise da semântica de palavras em distintos documentos é automática: essa é a diferença principal que existe entre o LSI e os outros modelos existentes." (p. 31).

c) Redes Neurais

Do ponto de vista conceitual as redes neurais podem ser vistas como uma estratégia ou técnica inspirada no funcionamento do cérebro, em que neurônios artificiais, conectados em rede, são capazes de aprender e de generalizar. Do ponto de vista matemático, é uma técnica de aproximação de funções por regressão não linear. Nesse modelo elas são usadas para realizar o casamento entre consultas e os documentos armazenados no acervo de um SRI. Cada consulta deflagra um processo que dispara sinais ativadores dos termos índices e esses, por sua vez, propagam os sinais aos documentos relacionados. Então, os documentos relacionados retornam os sinais a novos termos índices, em sucessivas interações. O conjunto resposta é definido através desse processo e pode conter documentos que não compartilham nenhum termo-índice com a consulta, mas que porém tenham sido ativados durante o processo, conforme BAEZA-YATES & RIBEIRO-NETO (1999) analisado por SOUZA (2006).

5.6.1.3 Modelo Probabilístico

É uma visão importada do campo das Probabilidades, uma área da Estatística. Constitui-se em um dos modelos clássicos de recuperação de informação baseado na interpretação probabilística da relevância dum documento para uma dada interrogação. O modelo parte da premissa de que existe um conjunto ideal de documentos que satisfaz a cada uma das consultas ao sistemas e que esse conjunto pode ser recuperado. Nele os termos indexados dos documentos e das consultas não possuem pesos pré-definidos. A atividade de ordenação dos documentos é realizada "pesando-se" dinamicamente os termos da consulta relativamente aos documentos. É baseado no princípio da ordenação

probabilística *(Probability Ranking Principl*e). De acordo com RIJSBERGEN (1979), nesse modelo, buscamos determinar a probabilidade de um documento *específico* ser ou não relevante para uma consulta específica. Tal informação pode ser obtida assumindo-se que a distribuição de termos na coleção seja capaz de informar a relevância provável para um documento qualquer da coleção. De acordo com WIVES (2002), nesse modelo a função de similaridade pode se aproveitar das informações estatísticas de distribuição dos termos contidos no índice. Assim sendo, determinados parâmetros podem ser ajustados de acordo com a coleção em questão, obtendo com isso resultados mais relevantes. Conforme SOUZA (2006):

> *"Através de tentativa inicial com um conjunto de documentos (para a qual se podem utilizar técnicas de outros modelos, como o vetorial) e do feedback do usuário em sucessivas interações, busca-se aproximar cada vez mais deste conjunto ideal, por meio de análise dos documentos considerados pertinentes pelo usuário. O valor desse modelo está em considerar a interação contínua com o usuário como um caminho para refinar o resultado continuamente".*

Esse modelo pode ser desdobrado e aperfeiçoado pelas redes de inferência e pelas redes de crença, ambas vinculadas à concepção de redes *bayesianas.*

a) Redes de Inferência (ou Redes *Bayesianas*) [42]

As redes bayesianas são estruturas para representar as dependências entre variáveis e fornecer uma especificação concisa de qualquer distribuição ou probabilidade conjunta total, conforme RUSSEL & NORVIG (2004), que alega que tal rede é um grafo orientado em que cada nó é identificado com informações de probabilidade quantitativa. De acordo com esse autor, a especificação completa de uma rede bayesiana é dada por:

> *a) Um conjunto de variáveis aleatórias constitui os nós da rede e podem ser discretas ou contínuas;*
> *b) Um conjunto de vínculos orientados ou seta conecta pares de nós. Se houver uma seta do nó X até o nó Y, X será denominado pai de Y;*
> *c) Cada nó Xi tem uma distribuição de probabilidade condicional P que quantifica o efeito dos pais sobre o nó.*

[42] Outros nomes também são usados, tais como redes probabilísticas, redes causais, mapas de conhecimento. Em estatística, a expressão modelo gráfico s refere a uma classe um pouco mais ampla que inclui as redes bayesianas. Essas redes são bem descritas em RUSSEL e NORVIG, 2004, cap. 14.

> *A topologia da rede – o conjunto de nós e vínculos – especifica os relacionamentos de independência condicional que são validados no domínio. O significado intuitivo de uma seta em uma rede construída corretamente é de que X tem uma influência direta sobre Y. Normalmente é fácil para um especialista em domínios descobrir quais são as influencias existentes no domínio, embora não seja fácil especificar as probabilidades em si* (p. 480).

Um exemplo de rede bayesiana simples pode ser apresentado na figura 22 a seguir. Nela, há uma variável t que é completamente independente das demais. Poderia ser o tempo, por exemplo. Outras três variáveis são articuladas, sendo que y e z são condicionalmente independentes, dada a variável x.

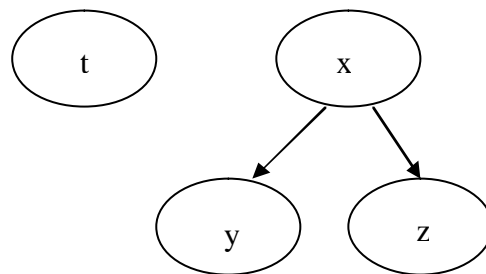



Figura 22: Representação de uma rede *bayesiana* simples
Baseado em RUSSEL & NORVIG (2004, p. 480)

Uma rede bayesiana é uma representação bem desenvolvida para o conhecimento incerto. Elas fornecem um modo conciso de representar relacionamentos de independência condicional no domínio. Portanto, para a recuperação da informação, nas redes *bayesianas*, ou de inferência (*inference network model*), são associadas variáveis aleatórias ao evento do atendimento de uma consulta específica por um documento específico. Uma variável aleatória associada com um documento específico representa o evento de observar o documento, sendo que esse tipo de variável pode ser modificada conforme os eventos futuros, de forma a estabelecer relacionamentos baseados nos eventos observados. Este modelo leva em conta o *feedback* de relevância do usuário.

b) Redes de Crença

As redes de crença (*belief network model*) são similares ou um subconjunto das redes inferência. Nelas, os documentos e as consultas são definidos como subconjuntos de um espaço de conceitos. A cada documento, associa-se a probabilidade de que o mesmo cubra os conceitos presentes no espaço de conceitos. Cada consulta é mapeada no espaço de conceitos, que por sua vez, está conectado ao espaço de documentos.

5.6.2 Modelos Estruturados de Recuperação em Textos

Os modelos clássicos respondem a consultas, buscando em uma estrutura de dados que representa o conteúdo dos documentos de uma coleção, unicamente como lista de termos significativos. Quando uma consulta é baseada numa palavra-chave, nós assumimos que existe uniformidade nos documentos e que uma palavra qualquer tem a mesma importância em qualquer ponto onde apareça no documento. Assim sendo, não se diferencia se a palavra aparece no título ou no corpo do texto. Termos que apareçam no título, subtítulo ou indicando uma parte específica em um formulário que contém vários diferentes campos possuem a tendência de ter maior importância. Os modelos que combinam as informações contidas no texto com as estruturas dos documentos onde se encontram são chamados de modelos estruturados de recuperação em textos.

Um modelo de recuperação de documentos estruturados utiliza a estrutura dos mesmos, buscando melhorar o desempenho e proporcionar ao usuário serviços alternativos, como por exemplo, uso de memória visual, recuperação de elementos multimídia, maior precisão no campo da consulta, dentre outros aspectos. Para esses modelos, temos alguns conceitos ou elementos fundamentais, quais sejam: os pontos de unificação, as regiões do texto e os nós. Um ponto de unificação ("*match point*") é a posição de uma seqüência de palavras que possa satisfazer ou casar com uma consulta de um usuário. Em uma consulta simples, como "*cães pastores na Escócia*" temos três pontos de unificação. Região significa um uma região contígua de texto. Já o conceito de nó se refere a um componente estrutural do documento, tais como um capitulo, seção ou subseção, etc. (BAEZA-YATES & RIBEIRO-NETO (1999, p.63).

Ao considerar a descrição da parte da estrutura de um documento é possível gerar um grafo representando-a e sobre a qual se possa navegar, sendo que ela será útil para responder certas consultas de alguns tipos, como por exemplo:

- Sobre a estrutura: quais são as seções do segundo capítulo;
- Sobre metadados: documentos da Eletronuclear editados em 2005;
- Sobre o conteúdo: que possuem a palavra energia nos títulos das seções;
- Sobre elementos multimídia: imagens cercando parágrafo que contenham a palavra "usina".

Enquadram-se na categoria de modelos estruturados o modelo das listas não sobrepostas e o modelo da proximidade de nós.

5.6.2.1 Listas Não-Sobrepostas

BURKOWSKI (1992) propõe dividir o texto de cada documento em regiões que não estão sobrepostas e juntá-las em uma lista. Existem muitas maneiras de dividir o texto em regiões não sobrepostas. Diferentes formas de dividir o texto em regiões não-sobrepostas levam a geração de listas múltiplas, por exemplo: uma lista para capítulos, uma lista para seções e outra para subseções, conforme figura abaixo:

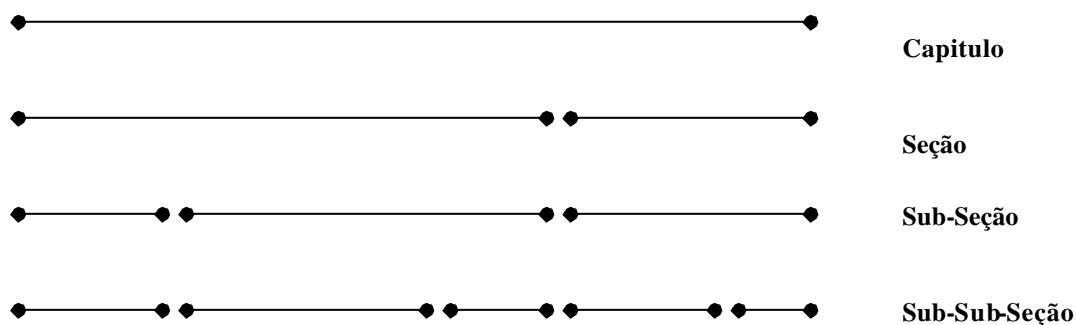



Figura 23: Representação da estrutura em um texto de um documento através de três listas (planos) de indexação.
FONTE: BAEZA-YATES & RIBEIRO-NETO (1999 p. 63)

Para permitir a busca por regiões de texto, especialmente por termos de indexação, constrói-se um índice, mediante um arquivo invertido,[43] em que há uma entrada por cada tipo de componente estrutural e associada a cada entrada há uma lista das regiões que representam nós deste tipo. Cada lista pode ser mesclada com o índice correspondentes dos diferentes termos. A partir do momento que as regiões de texto estão não-sobrepostas, três questões podem ser respondidas:

- selecione uma região que contenha uma dada palavra;
- selecione uma região A que não contenha uma região B (A e B pertencem a listas distintas);
- selecione uma região que não esteja contida em qualquer outra região;

Esse modelo é simples e permite uma implantação eficiente, embora o tipo de questões que podem ser respondidas seja, até certo ponto, limitado.

5.6.2.2 Proximidade de Nós

Esse modelo foi introduzido por NAVARRO E BAEZA- YATES (1997). É um modelo que visa a recuperação de documentos, através de uma estrutura hierárquica de índices. Então, a proposta central é definir um índice hierárquico estrito sobre o texto. Isto enriquece o modelo anterior. Esse modelo permite a definição de estruturas de indexação hierárquicas e independentes sobre um mesmo documento. Cada uma dessas estruturas é uma hierarquia estrita composta por nós, que podem ser capítulos, seções subseções, parágrafos e linhas. A cada nó é associada uma região de texto. Assim duas hierarquias diferentes podem fazer referência a regiões de texto que se sobrepõe.

---

[43] Tipo de mecanismo utilizado para localizar um dado termo em um texto. Um arquivo invertido é um tipo de arquivo indexado, conforme SALTON & MCGILL (1983). Em um arquivo indexado define-se uma estrutura complementar de índices permitindo-se que os documentos recebam partições, facilitando assim uma recuperação ulterior. Um arquivo invertido é normalmente composto de um vocabulário – um arranjo contendo todas as palavras distintas do texto – e para cada palavra do vocabulário, uma lista de todos osdocumentos (identificados por números de documentos mos quais aquela palavra ocorre. Para entender melhor, considere um conjunto de documentos. A cada documento é atribuído um conjunto de palavras-chave ou atributos. Um *arquivo invertido* é constituído de uma lista ordenada (ou índice) de palavras-chave (atributos), onde cada palavra-chave tem uma lista de apontadores para os documentos que contêm aquela palavra-chave. Este é o tipo de índice utilizado pela maioria dos sistemas para recuperação em arquivos constituídos de texto (ZIVIANI, Nívio. UFMG/DCC. in: http://homepages.dcc.ufmg.br/~nivio/cursos/ri05/tp1/ri05tp1.html . Acessado em 12 dez 2006.

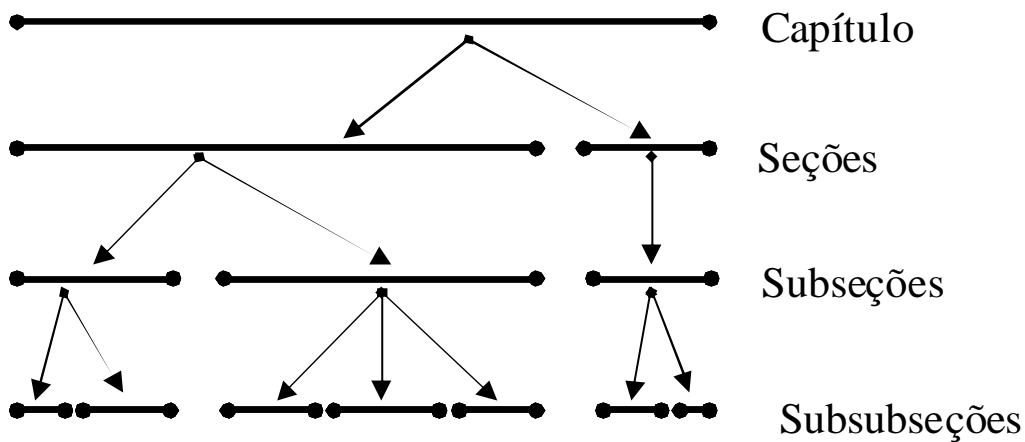



Figura 24: Representação da estrutura de um documento mediante o modelo de proximidade de nós.

Fonte: BAEZA-YATES & RIBEIRO-NETO (1999 **p.64)**

5.6.2.3 Recuperação por Passagens

A recuperação por partes de documentos é antiga e seu princípio bastante simples: dentro dos documentos textuais as passagens são seqüências textuais de palavras consecutivas que apresentam certa homogeneidade em sua temática e que, portanto, podem ser relevantes para um usuário, entendendo-se aqui palavras consecutivas como uma seqüência de palavras segundo uma ordem de leitura do documento desde a primeira até a última página. Uma vez tratadas, as passagens são consideradas documentos separados e se indexam e recuperam através dos modelos clássicos já descritos. Esta abordagem é mais útil para documentos longos ou com estrutura complexa.

Conforme CALLAN (1994) a recuperação por passagens procura evidências de relevância em vários níveis do documento. A busca consiste em segmentar o texto e comparar a consulta com cada segmento. Uma divisão do documento pode ser feita com base no discurso (parágrafos, seções, frases), na semântica (em conteúdo ou assunto) ou por janela (conforme um determinado número de palavras próximas).

5.6.3 Modelos por Aglomerados e Contextual

No detalhamento dos modeles de recuperação textual WIVES (2002) enumera os seguintes:

- *booleano*;
- espaço vetorial;
- probabilístico;
- difuso;
- busca direta;
- aglomerados;
- lógico;
- contextual.

Comentaremos os modelos por Aglomerados e Contextual.

5.6.3.1 Modelo Por Aglomerados

De acordo com WIVES (2002, p.43-44) o conceito básico envolvido nessa abordagem consiste em identificar documentos de conteúdo similar (que tratem de assuntos similares) e armazená-los ou indexa-los em um mesmo grupo ou *aglomerado (cluste*r). A identificação de documentos similares em conteúdo dá-se pela quantidade de palavras similares e freqüentes que eles contêm. Quando o usuário especifica sua consulta e essa é remetida ao sistema, que, por sua vez, identifica um documento relevante (e isso pode ser feito através de técnicas de "casamento" tradicionais ou através de técnicas específicas para grupos) e retorna para o usuário todos os documentos pertencentes ao mesmo grupo. Em teoria, os documentos pertencentes a um mesmo grupo também são relevantes à consulta, uma vez que diz que objetos semelhantes e relevantes a um mesmo assunto tendem a permanecer em um mesmo grupo e possuem atributos em comum (hipótese de agrupamento).

Considerando que buscas tradicionais ignoram o co-relacionamento entre documentos, o modelo de aglomerados tende a aumentar a qualidade dos resultados

(pois retorna todo o grupo coeso e relevante à consulta) e o tempo de processamento (já que os grupos de documentos tendem a ser armazenados em um mesmo bloco do dispositivo de armazenamento). Se o agrupamento for realizado de forma hierárquica é possível oferecer um sistema de busca por navegação, aonde o usuário vai selecionando os ramos que considera mais adequados até encontrar o grupo de documentos mais relevantes. Um dos maiores problemas desse modelo é justamente identificar os grupos de documentos mais coesos e mantê-los assim durante a utilização do sistema. Todo documento inserido ou modificado deve ser re-analisado a fim de ser colocado no grupo correto. Não há muitas informações sobre *clustering* como um modelo, mas sim como técnica de descoberta de conhecimento

5.6.3.2 Modelo Contextual

Esse modelo pode ser fortemente identificado ou acoplado à Indexação Semântica Latente, anteriormente descrita. Ele também é descrito por WIVES (2002, p.45-47), com base em LOH (1997) a partir do qual é apresentada a descrição a seguir. A maior parte dos modelos de SRIs considera a presença dos termos nos documentos e fazem o "casamento" entre um documento e a consulta exclusivamente se as palavras contidas no documento forem exatamente iguais (casando os padrões similares) às palavras especificadas na consulta. Assim sendo, os documentos que possuem as palavras identificadas na consulta são considerados relevantes e os que não as possuem (mesmo que os termos tenham o mesmo sentido) são considerados irrelevantes por possuírem uma *morfologia* diferente. Esse tipo de "casamento" é muito restritivo, pois, como já salientado, a linguagem natural possui ambigüidade e incerteza inerentes, causando problemas de sinonímia (onde vários termos podem denotar um mesmo objeto) e polissemia (onde um termo possui vários significados). Esses são os *problemas do vocabulário* (ou problemas da diferença de vocabulário). Com isso, se um documento trata do assunto especificado pela necessidade do usuário, mas seu autor não utiliza os mesmos termos que o usuário, esse documento não é considerado relevante.

O modelo contextual (ou conceitual) é desenvolvido a partir do princípio de que todo documento possui um contexto, pois a pessoa que escreve um texto o faz desenvolvendo um assunto específico e utiliza frases interconectadas ou encadeadas que

fazem sentido dentro assunto (o contexto). A consulta do usuário também possui um contexto que é definido por sua necessidade de informação. Uma vez identificado o contexto dessa necessidade de informação e os contextos dos documentos de uma coleção (base de documentos), o processo de recuperação e de identificação de informações relevantes pode ser feito ao nível de contextos e não mais ao nível de palavras isoladas. Espera-se com isso que os resultados em relação à relevância dos documentos retornados sejam melhores. A princípio, tal idéia parece prática e fácil, porém, identificar e modelar os contextos dos documentos não é uma tarefa trivial. Os processos de cognição humana ainda não são completamente compreendidos e, portanto, não é possível saber que elementos são necessários para modelar um contexto. Atualmente isso é feito selecionando algumas palavras que em conjunto (e estando correlacionadas) podem definir (dar uma idéia de) esse contexto. Como cada palavra pode estar presente em mais de um contexto, deve haver um grau (peso) indicando quanto uma palavra é relevante (importante) em cada contexto. Esse conjunto de palavras é então utilizado para representar o contexto (LOH, 1997). Os documentos são então indexados de acordo com os contextos existentes e definidos. Isso é feito através de uma espécie de classificação onde as características (palavras ou termos) que descrevem determinado contexto são localizadas nos documentos. Cada característica encontrada ativa seu contexto correspondente (palavras com graus elevados de relevância podem definir um contexto pelo simples fato de aparecerem em um documento). O valor de relevância (importância) dessa palavra é adicionado ao grau de pertinência do documento aos contextos que ela representa. Logo, quanto mais palavras localizadas (ativas) um contexto possuir, maior é o grau de relação do documento com esse contexto. Nota-se com isso que um documento pode pertencer a mais de um contexto com graus diferentes de relevância (relação ou pertinência).

A identificação do contexto da consulta do usuário também é feita da mesma forma: as palavras existentes ativam o contexto mais relevante. Dependendo do sistema o usuário pode “navegar” pelos contextos existentes e selecionar um para a busca, encontrando assim todos os documentos que estejam relacionados com o contexto escolhido (LOH 1997).

Esse modelo não elimina o problema do vocabulário, mas pode minimizá-lo se o conjunto de palavras utilizado na descrição dos contextos for bem escolhido. Várias

palavras podem ser utilizadas nessa descrição. Porém, muitas delas certamente são encontradas em vários contextos. Devem ser escolhidas aquelas que caracterizam bem cada contexto sem que indiquem ou apareçam em outros (ou muitos outros) contextos. Ou seja, as palavras devem ter um alto grau de discriminação. Isso induz que os contextos sejam descritos manualmente por uma pessoa que conheça muito bem o assunto (contexto) que está descrevendo. O ideal é que essa pessoa seja um especialista na área (contexto) em questão. Essa pessoa deve selecionar os termos mais relevantes do contexto, adicionando também sinônimos e palavras específicas. Pelo fato de muitas dessas palavras poderem aparecer em mais de uma descrição de contexto, além do fato de umas poderem ser mais relevantes (descritivas ou discriminantes) do que outras, um grau (peso) de relevância (ou importância) deve ser atribuído a cada palavra. Essa descrição manual de contextos pode ser auxiliada (ou substituída, dependendo do caso) por um processo automatizado de identificação de palavras relevantes. Existe uma série de técnicas desenvolvidas com o propósito de identificar palavras importantes e correlações entre palavras (SALTON & MACGILL, 1983), que podem ser utilizadas. Dicionários de sinônimos, thesaurus e redes semânticas também podem ser utilizadas nesse processo.

Um dos maiores problemas desse modelo está no fato dos descritores poderem ser elaborados incorretamente, o que ocasionaria uma busca incorreta onde os documentos retornados provavelmente seriam irrelevantes para a necessidade do usuário. Ou seja, a descrição dos contextos deve ser elaborada cuidadosamente para que a recuperação contextual funcione de forma a oferecer resultados relevantes e coerentes.

5.7 Análise das Taxonomias dos Modelos e Perspectivas

Existem limites em modelos e métodos da recuperação da informação que poderíamos chamar de tradicional, ou seja: quando se trabalha com palavras de um documento de modo isolado. Nesse caso, somente se recuperam palavras por semelhança literal, levando assim à perda da informação existente no conjunto de palavras e de documentos. Precisamos considerar que os documentos que precisamos recuperar não precisam ter obrigatoriamente os mesmos termos utilizados para se especificar uma consulta. Daí deriva a necessidade do esforço por um processo de recuperação que vai se revestindo, na verdade em processo de extração de informação

(ou de conhecimento), considerado as interações entre as palavras, seus conjuntos e mesmo as estruturas de diferentes documentos em uma base. Caminha-se na direção de um esforço por tratamento semântico dos documentos para responder às necessidades de informação dos usuários. Esse é um esforço que demanda complexidade e uso de diferentes recursos metodológicos e computacionais cada vez mais sofisticados e muitas vezes mesclados. Novos métodos, combinados com tradicionais estratégias, produzem novos efeitos de eficiência para o usuário.

Obviamente as taxonomias apresentadas sobre modelos de recuperação da informação podem ser questionadas sob diferentes prismas. A recuperação da informação tornou-se hoje um campo amplo, em que se entrelaçam diferentes teorias, visões e conceitos, bem como tecnologias de informação. Por exemplo, a inclusão de lógica nebulosa (*fuzzy*) e redes neurais pode ser alvo de reflexão, uma vez que essas abordagens são tratadas no campo da prospecção de conhecimento, vinculado à inteligência artificial. Portanto, essa taxonomia mescla aspectos e visões tradicionalmente consideradas no campo da recuperação da informação com outras oriundas da Inteligência Artificial. Uma análise que podemos fazer para isso é o fato de que os SRIs tiveram que estender suas habilidades de lidar com a informação para responder ás crescentes necessidades dos usuários, cada vez mais complexas e exigentes e termos de cotejamento, organização e apresentação dessas informações. Os SRIs já se diferenciaram de modo consolidado dos sistemas baseados em recuperação de dados, mas é necessário hoje que expandam ainda mais seus recursos e habilidades. Tal expansão vem acontecendo de forma frutuosa quando caminha em direção à melhoria de técnicas matemáticas e estatísticas e, mais especialmente no rumo dos sistemas de descoberta de conhecimento, notadamente os que possuem bases na inteligência artificial com as perspectivas de processamento de linguagem natural. Embora já citados preliminarmente ao longo desse capítulo, em diversos pontos, aspectos relativos à descoberta de conhecimento serão abordados no capítulo 6.

6 SISTEMAS INTELIGENTES E PROSPECÇÃO DE CONHECIMENTO

6.1 Considerações Iniciais

Cada vez mais os SRIs incorporam novas funcionalidades de forma a melhor responder às necessidades dos usuários. Os Sistemas de Recuperação Textuais, sobretudo após a consolidação da Internet e do grande crescimento de bases textuais nas organizações, cresceram em complexidade e disponibilidade de recursos para os usuários, cada vez mais sufocados pelo turbilhão de documentos disponíveis, com demandas sofisticadas por informações muitas vezes articuladas, porém sem condições de dar saltos superiores aos permitidos por tradicionais motores de busca. Nas organizações, devido a um ambiente cada vez mais complexo e competitivo, é grande a necessidade de informação. Essas necessidades, de usuários individuais e de organizações, para serem devidamente respondidas, passaram a requerer um tratamento semântico dos conjuntos de documentos e um processamento computacional cada vez mais complexo, o que passou a demandar novos recursos como os providos pela Inteligência Artificial, com seus sistemas inteligentes voltados para o que se denomina descoberta de conhecimento.

A idéia de uso de sistemas "inteligentes" está vinculada à pretensão do ser humano fazer com que recursos tecnológicos computacionais possam realizar tarefas análogas à que o ser humano realiza, procurando reproduzir, até certo ponto, as características comportamentais inteligentes que o homem possui. É, obviamente, uma abstração, uma vez que o conceito de inteligência é complexo (e não será aqui explorado, mais tangenciado). O ser humano é capaz, intelectualmente, de representar os objetos do mundo de forma abstrata, produzir e manipular conceitos, trabalhar com fatos que guarda na memória, cotejar esses fatos e conceitos e experiências entre si. Através da cognição, o homem absorve a e analisa fatos, que se transformam e registros e gradativamente vai definindo símbolos que serão manipulados pela linguagem e pelo pensamento. Representa encadeamentos por uma visão lógica dos fenômenos e trabalha com associações, generalizações e a criatividade. Em seu confronto com a realidade, conforme o processo de "*equilibração majorante*" (PIAGET, 1976), descrito no capitulo 4, vai estabelecendo formas de se relacionar com o mundo de forma dinâmica, adaptativa e evolutiva, sempre inovando e apresentando a possibilidade de mudanças

continuas, seja pela realização de novas experiências seja pela capacidade de, continuamente elaborar e manejar novos conceitos. Enfim, o ser humano tem um comportamento que, com tais características pode ser definido como inteligente, utilizando conhecimento e raciocínio. O conhecimento está articulado com a idéia de dados e informações que, de acordo com REZENDE (2003), conforme já apresentado no capitulo 4, diz respeito à habilidade de criar um modelo mental que descreva o objeto e indique as ações a implementar e as decisões a tomar. A compreensão, a análise e síntese, necessárias para a tomada de decisões inteligentes, são realizadas a partir do nível do conhecimento. Conforme essa autora, a geração de conhecimento é resultado de:

> "*um processo no qual uma informação é comparada com outra e combinada em muitas ligações (hiperconexões) úteis e com significado. Isso implica que o conhecimento é dependente de nossos valores e nossa experiência e sujeito às leis universalmente aceitas. (...) Uma decisão é um uso explícito de conhecimento. O conhecimento pode ser representado como uma combinação de estruturas de dados e procedimentos interpretativos que levam a um comportamento conhecido. Esse comportamento fornece informações a um sistema*[44] *que pode, então, planejar e decidir* (p.5-6)".

Ainda de acordo com REZENDE (2003) o conhecimento pode ser:

- Declarativo – descritivo e genérico sobre os fatos e eventos (o que é)
- Procedural – prescritivo e resultado de encadeamentos (como funciona isto ou aquilo)
- Senso comum – composto de conhecimento declarativo e procedural e realizando uma avaliação ou julgamento do que é "certo ou errado"
- Heurístico – conhecimento único para cada indivíduo ou grupos bem específicos, que dificilmente pode ser obtido em alguma fonte, envolvendo avaliação sistemática e uso de regras heurísticas. Está fortemente associado á idéia de conhecimento tácito e tem valor importante em inteligência artificial.

Ao solucionar problemas e interagir com o mundo o ser humano pode utilizar cada um desses diferentes tipos de conhecimento, a partir de diferentes combinações de informações. Tanto a análise lógica, baseada num conhecimento explícito expresso por

[44] Nesse caso, já considerando a idéia de uma possível aplicação de sistema informatizado.

dados objetivados em relatórios ou sistemas informatizados, quanto uma análise subjetiva ou heurística, baseada na experiência e na intuição podem ser aplicados.

6.2 Conceito de Inteligência Artificial e de Sistemas Inteligentes

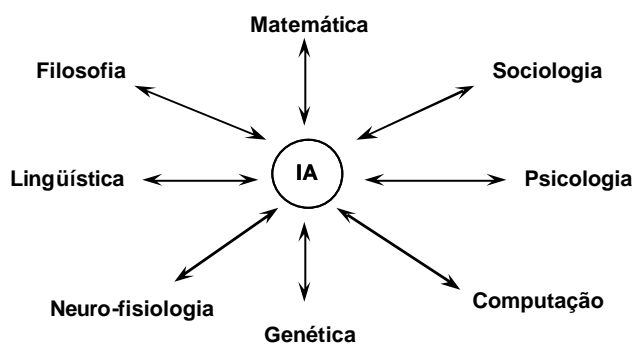
Quando dizemos que os sistemas de informação computadorizados são inteligentes e estão apoiados em conhecimento estamos na verdade fazendo uma analogia com tal comportamento humano, uma vez que, por mais avançados que sejam esses dispositivos, eles possuem limites explícitos quando confrontados com o ainda indefinido limite do pensamento humano. O esforço de transferir para sistemas computadorizados os processos humanos de inteligência, envolvendo as relações com o conhecimento e o raciocínio deu origem a um campo denominado inteligência artificial. Esse campo é vasto e polêmico, comportando muitas abordagens e definições. Há também muitas críticas a ele. De qualquer forma, ele apresentou um significativo desenvolvimento, a partir de 1950. Foi fortemente questionamento quanto a seus resultados nos anos 60, tendo um período de refluxo, mas a partir dos anos 80 assumiu uma abordagem e uma postura mais moderna e condizente com as necessidades de pessoas e organizações. A redução dos custos de processamento de dados e por outro lado a elevação da capacidade computacional deste, criou novas perspectivas para esse campo do conhecimento, hoje muito mais inserido na resolução de problemas e suporte à tomada de decisões para o ser humano.

De acordo com TURBAN (2003), os softwares convencionais são baseados em algoritmos[45] e no processamento numérico. Os programas baseados em inteligência artificial vão além da computação convencional ao incluir a heurística ou normas práticas que expressam o conhecimento. Um sistema inteligente usa também o processamento simbólico do conhecimento, sendo que, em inteligência artificial, um símbolo pode ser uma letra, palavra ou número que representa objetos, processos e suas relações. Os objetos podem ser pessoas, coisas, idéias, conceitos eventos ou declarações de ocorrências. Usando símbolos, é possível criar uma base de conhecimento de fatos e conceitos e das relações existentes entre eles.

---

[45] Fórmulas matemáticas ou procedimentos seqüenciais que resultam em uma solução. Um algoritmo é convertido em um programa que informa ao computador as operações exa tas que devem ser executadas para solucionar problemas. TURBAN (2003, p. 403).

Adotando a visão de RUSSEL & NORVIG (2004) [46] podemos dizer que a inteligência está relacionada principalmente a uma ação racional. Então, no caso ideal, um agente inteligente adota a melhor ação possível em uma situação, em resposta a qualquer seqüência de percepções. [47] Um agente é algo que percebe e age em um ambiente. É um elemento capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por meio de atuadores. A inteligência artificial é, para tais autores, uma ciência de projeto de agentes inteligentes, ou seja: agentes bem sucedidos para a função que se propõe e que envolvem a intenção de simular ou atuar de forma análoga aos processos cognitivos e racionais humanos. É um campo de conhecimento que estuda a perspectiva e as possibilidades de automação de comportamento inteligente. O comportamento inteligente dos agentes em um determinado ambiente deve incluir as capacidades de percepção, raciocínio, aprendizado, comunicação, ação e planejamento. Um sistema baseado em abordagens da inteligência artificial deve ser capaz de adaptar-se a novas situações, raciocinar, entender relações entre fatos, descobrir significados, reconhecer elementos verdadeiros e falsos e aprender com base em sua experiência acumulada.

A inteligência artificial foi fortemente influenciada por vários outros campos do conhecimento humano, conforme representado no esquema abaixo e cada um desses campos para que ela possua a sua atual configuração.

[46] Abordagem desenvolvida nos capítulos 1 e 2 da obra dos autores e referencia.

[47] Para os autores, percepção são as entradas perceptivas do agente em qualquer momento dado e seqüência de percepções é a história completa de tudo o que o agente já percebeu. Em geral, a escolha de ação de ação de um agente em qualquer instante dado pode depender da seqüência inteira de percepções observadas (por ele) até o momento. Se pudermos especificar a escolha de ação do agente para toda a seqüência de percepções possível, então teremos dito quase tudo o que existe a dizer sobre o agente. Em termos matemáticos, o comportamento do agente é descrito pela função de agente, que mapeia qualquer seqüência de percepções específicas para uma ação. RUSSEL & NORVIG, 2004, p. 34-35.

Conforme TURBAN (2003), apesar dos esforços, a inteligência artificial ainda se encontra longe de apresentar uma inteligência significativa, embora estejam ajudando muito em tarefas que requerem inteligência humana para obter uma melhoria significativa de produtividade, qualidade e ciclo de produção. Nos últimos anos a inteligência artificial evoluiu muito e suas aplicações hoje envolvem:

- Matemática: demonstração de teoremas, resolução simbólica de equações, geometria, etc.
- Pesquisa operacional: otimização e busca heurística em geral
- Jogos diversos.
- Processamento de linguagem natural: tradução automática, verificadores ortográficos e sintáticos, análise semântica, dentre outros.
- Sistemas tutores: modelagem do aluno, escolha de estratégias pedagógicas, etc.
- Simular a percepção (visão, tato, audição, olfato, paladar) em aplicações.
- Robótica (software e hardware): manipulação, navegação, monitoramento, etc.

São diferentes os paradigmas utilizados para desenvolvimento dos agentes inteligentes pela inteligência artificial, conforme podemos sintetizar abaixo, com expressão sumária de alguns exemplos de modelos.

- Simbólico: metáfora lingüística (Ex. sistemas de produção)
- Conexionista: metáfora cerebral (Ex. redes neurais)
- Evolucionista: metáfora da natureza (Ex. algoritmos genéticos, vida artificial)
- Estatístico/Probabilístico (Ex. Redes Bayesianas, sistemas difusos)

Diante do exposto, e ampliando a visão colocada em REZENDE (2003) podemos dizer que os sistemas inteligentes são agentes ou conjuntos de agentes atuando de forma harmonizada, que possuem capacidade de usar conhecimento na resolução de problemas e realização de determinadas tarefas. Podem aproveitar associações e inferências para trabalhar com problemas complexos que sejam parecidos com problemas reais. Eles podem armazenar, processar e recuperar grandes conjuntos de informações e proceder múltiplos cotejamentos, explicitando certos resultados que

teríamos maiores dificuldades em processar humanamente. Deve ser capaz também de analisar o contexto em que se encontra e a ele se adaptar. Conforme a autora:

> "*o comportamento inteligente de um sistema é resultado de múltiplas e encadeadas decisões. A escolha da decisão, ou controle da decisão é baseada em critérios de desempenho, duração e risco. O controle de decisão é um processo pelo qual as soluções de um problema e as tomadas de decisão são seqüenciadas, sincronizadas, inter-relacionadas e direcionadas para fornecer o comportamento orientado ao objetivo de um sistema (p.7-8)*".

Dentro dos paradigmas acima citados, e também sob inspiração de REZENDE (2003), podemos dizer que os sistemas inteligentes podem ser desenvolvidos e implantados com base em técnicas e metodologias as quais podem ser aplicadas isolada ou conjuntamente para auxiliar o processo decisório e a resposta às necessidades de informação dos usuários. Essas abordagens, técnicas e metodologias incluem a aquisição de conhecimento, processos *OLAP* de *data warehousing*[48], o aprendizado de máquina, a indução de regras e árvores de decisão, as redes neurais, o raciocínio baseado em casos, a lógica nebulosa, a computação evolutiva, o processamento da linguagem natural, a descoberta de conhecimento em base de dados (mineração de dados) e a descoberta de conhecimento em textos (mineração de textos) dentre outros.

6.3 Visão Geral de Abordagens Técnicas Para Sistemas Inteligentes

Dentre as estratégias tecnológicas emergentes que são utilizadas no desenvolvimento de sistemas inteligentes podemos especificar sumariamente[49] as seguintes, com um sumário de sua descrição[50].

6.3.1 Aprendizado de Máquina

A idéia de aprendizagem, em inteligência artificial significa que um agente inteligente de software observa suas interações com o meio em que se encontra e com

---

[48] *OLAP* e *data warehousing* serão tratados mais detalhadamente no próximo capítulo.

[49] Considerando a amplitude do assunto, não é objetivo dessa tese aprofundar estudos relativos à inteligência artificial, mas apenas salientar sua relevância e papel atual, bem como algumas técnicas que realizam um tratamento mais elaborado de dados produzindo informações de alta representatividade para o processo de geração de conhecimento.

[50] Definições apoiadas na análise da obra de MEYER, M., BABER, R. e PFAFFENBERGER, B. (2000, 436-440)

seus próprios processos de tomada de decisão. O conceito de aprendizagem implica em que as percepções devem ser usadas não só para agir, mas também para melhorar a habilidade do agente para agir no futuro (RUSSEL & NORVIG, 2004). A aprendizagem de máquina é uma área da inteligência artificial que tem por objetivo desenvolver técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática (REZENDE, 2003). De acordo com essa autora:

> "*um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas por meio da solução bem-sucedida de problemas anteriores. Os diversos sistemas de aprendizagem de máquina possuem características particulares e comuns que possibilitam sua classificação quanto à linguagem de descrição, modo, paradigma e forma de aprendizado utilizado* (p. 90)".

O aprendizado de máquina procura levar os softwares a aprender com os dados que eles estudam, de tal modo que esses softwares tomem decisões diferentes baseados em características dos dados estudados, utilizando a estatística para os conceitos fundamentais e adicionando mais heurística avançada da inteligência artificial e algoritmos para atingir seus objetivos. (SERRA, 2002). A aprendizagem pode ser supervisionada (indutiva) ou não-supervisionada (probabilística, baseada, em geral, em métodos estatísticos).

A aprendizagem supervisionada envolve a aprendizagem de uma função a partir de dados de entradas e saídas. A aprendizagem supervisionada de uma função de valores discretos é chamada de classificação, enquanto a aprendizagem de uma função com valores contínuos é chamada de regressão. Uma aprendizagem indutiva envolve encontrar uma hipótese consistente que concorde com os exemplos. (RUSSEL & NORVIG, 2003). Nela, é fornecido um algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. O objetivo desse algoritmo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados, ou seja: exemplos que não tenham o rótulo da classe (REZENDE, 2003).

A aprendizagem não supervisionada envolve métodos estatísticos de aprendizagem, em que visualizamos a aprendizagem como uma forma de raciocínio

com incerteza a partir de observações. De acordo com RUSSEL & NORVIG (2004) tais métodos variam desde o cálculo simples de médias até a construção de modelos complexos, como redes bayesianas e, segundo alguns autores, também redes neurais[51], com aplicação ampla no mundo científico e organizacional.

REZENDE (2003)[52] apresenta uma taxonomia para os paradigmas de aprendizado, classificando os em:

- Simbólico – buscam aprender construindo representações simbólicas de um conceito através da análise de exemplos e contra-exemplos desse conceito. Tais representações simbólicas se expressam sob a forma de expressões lógicas, árvores de decisão, indução de regras ou redes semânticas.
- Estatístico – a idéia consiste em encontrar modelos estatísticos para aproximações dos conceitos tratados. Vários desses métodos são paramétricos, assumindo alguma forma de modelo e encontrando valores apropriados para os parâmetros desse modelo a partir dos dados.
- Baseados em Exemplos – implica em classificar um exemplo com base na consideração de outro exemplo similar cuja classe já seja conhecida, assumindo que o novo exemplo ficará nessa mesma classe, ou seja: classificar exemplos nunca vistos por meio de exemplos similares conhecidos. O raciocínio baseado e casos é uma técnica desse paradigma.
- Evolutivo – Baseado em algoritmos genéticos, descritos no tópico 6.3.2 dessa tese.
- Conexionista – Está baseado na abordagem das redes neurais, descrita no tópico 6.3.4 dessa tese.

6.3.2 Algorítmos Genéticos

Algoritmos Genéticos são algoritmos eficientes de busca probabilística, inspirados nos mecanismos da evolução biológica. Algoritmos Genéticos têm produzido excelentes soluções em problemas complexos de otimização que possuem grande

[51] Embora as redes neurais possam ser vistas dentro de um paradigma conexionista e não eminentemente estatístico.
[52] REZENDE (2003, p. 93-94)

número de parâmetros. Áreas onde esses algoritmos têm sido aplicados incluem: otimização de funções complexas, planejamento industrial, lay-out de circuitos, controle e otimização de distribuição de energia, otimização de roteamento e tráfego, etc.

O objetivo principal de um algoritmo genético é o de evoluir a partir de uma população de soluções para um determinado problema, tentando produzir novas gerações de soluções que sejam melhores que as anteriores. Esses algoritmos operam através de um ciclo simples que consiste dos seguintes passos: criação da população inicial, avaliação, seleção e reprodução, este último envolvendo a recombinação e mutação do material "genético" das soluções.

6.3.3 Lógica Difusa

A lógica difusa, ou nebulosa, fornece um mecanismo para computadores lidarem com informações imprecisas. A lógica difusa – *fuzzy* – foi projetada para manipular conceitos imprecisos como "pequeno", "grande", "jovem", "velho", "alto", ou "baixo", que não utilizam uma lógica binária do "branco ou preto", mas um esquema lógico que possa manipular "tons de cinza". Sistemas baseados na lógica nebulosa têm mostrado grande utilidade em uma variedade de operações de controle industrial e em tarefas de reconhecimento de padrões que se estendem desde reconhecimento de texto manuscrito, até a avaliação de crédito financeiro. Existe também um interesse crescente em se utilizar lógica nebulosa em sistemas especialistas para torná-los mais flexíveis.

Uma das principais potencialidades da lógica nebulosa, quando comparada com outros esquemas que tratam com dados imprecisos como redes neurais, é que suas bases de conhecimento, as quais estão no formato de regras de produção, são fáceis de examinar e entender. Este formato de regra também torna fácil a manutenção e a atualização da base de conhecimento.

6.3.4 Redes Neurais Artificiais

Redes neurais (*neural networks*), ou “neuronais” para alguns autores, são dispositivos computacionais, inspirados na funcionalidade dos neurônios biológicos, aplicados ao reconhecimento de padrões. A sua habilidade em formar mapeamentos

não-lineares tem tornado as redes neurais prósperas na modelagem e previsão de sistemas complexos.

Redes neurais são compostas por diversas unidades computacionais paralelas interconectadas. Cada uma dessas unidades efetua um pequeno número de operações simples e transmite seus resultados às unidades vizinhas. As redes neurais aprendem a reconhecer padrões através de um processo de treinamento baseado em vários exemplos diferentes. Elas são eficientes no reconhecimento de padrões em um conjunto de dados, quer estes representem imagens digitalizadas ou o comportamento de uma moeda no mercado de câmbio. Conforme TURBAN (2003) as redes neurais constituem um método diferente que usa a arquitetura que simula determinadas capacidades de processamento do cérebro humano. Os resultados são representações do conhecimento e processamento baseado no método paralelo em massa (execução de várias instruções de processamento simultaneamente). Um neurônio artificial recebe entradas análogas aos impulsos eletroquímicos que os neurônios biológicos recebem de outros neurônios. Os neurônios de uma rede neural recebem informações de outros neurônios ou de fontes externas, transformam as informações e as transmitem para outros neurônios ou para saídas externas. A figura 25 abaixo representa essa idéia.

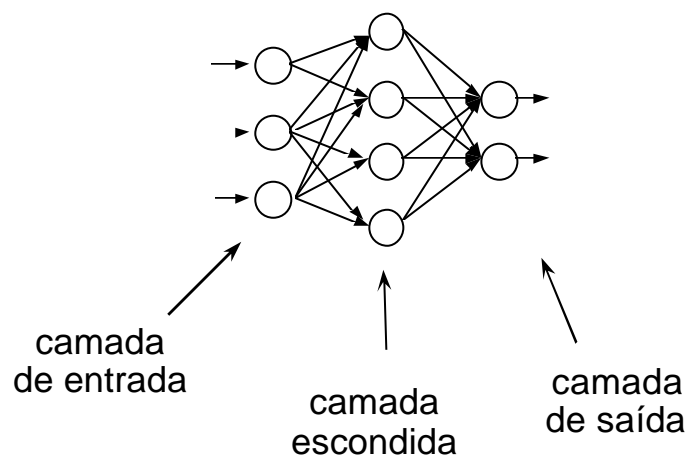



Figura 25: Representação Para Entendimento de uma Rede Neural

Redes neurais também são bastante efetivas no aprendizado de padrões a partir de dados não-lineares, incompletos, com ruídos e até compostos de exemplos contraditórios. A habilidade de manipular estes dados imprecisos faz com que as redes neurais sejam extremamente eficazes no processamento de informações financeiras e

em tarefas onde especialistas não estão à disposição ou um conjunto de regras claras não pode ser facilmente formulado. As redes neurais não são, portanto, apropriadas para as aplicações em que a explicação do raciocínio é fundamental.

6.3.5 Árvores de Decisão

Árvores de decisão e indução de regras são técnicas de aprendizagem de máquina simbólica. É uma técnica que, a partir de uma massa de dados, cria e organiza regras de classificação e decisão em formatos de diagramas de árvores, que irão classificar suas observações ou predizer resultados futuros. Por exemplo, se seus dados estiverem divididos em classes dicotômicas, tais como: bons e maus pagadores, clientes e não-clientes, infectados e não infectados, uma árvore de decisão pode ser construída para criar regras que classifiquem casos já existentes, ou casos novos, com precisão. Começa-se com um único grupo que reúne todos os casos em estudo. Na medida em que a árvore vai se expandindo, esta base é divida em nódulos que representam categorias das variáveis analisadas. Cada galho da árvore é formado por esses nódulos eu vão se abrindo em subgrupos mutuamente exclusivos. Cada nódulo e cada galho apresentam uma proporção de obtenção da resposta em estudo (BARBIERI, 2001)

6.3.6 Sistemas Híbridos

Recentemente tem havido considerável interesse no desenvolvimento de sistemas híbridos envolvendo técnicas computacionais inteligentes. Sistemas híbridos combinam técnicas de processamento simbólico e adaptativo, tais como sistemas especialistas, lógica nebulosa, redes neurais e algoritmos genéticos. Tais sistemas apresentam diferentes características quanto à funcionalidade e à arquitetura.

Defensores de ambos, processamento simbólico e adaptativo, argumentam que suas técnicas isoladamente são as mais adequadas para modelar e resolver problemas complexos do mundo real. Um exame em detalhe dessas classes de processamento e de suas aplicações revela, no entanto, que ambos os procedimentos apresentam uma combinação de vantagens e limitações e que a integração de diferentes técnicas pode ajudar a superar seus pontos fracos. Na verdade, muitos dos problemas complexos do mundo real requerem a ação combinada dos procedimentos complementares.

6.3.7 Sistemas Fórmicos (Ant *Systems*)

Conforme DORIGO & STÜTZLE (2006), o complexo comportamento social das formigas está atraindo muitos estudos pela ciência. A Ciência da Computação têm buscado nos padrões desse comportamento inspiração para prover modelos de resolução de problemas difíceis de otimização combinatória. Sistemas fórmicos, ou *ant systems* são um novo método heurístico para resolver problemas complexos que exigem otimização combinatória. A otimização combinatória pode ser entendida nos seguintes termos:

- Partindo de um conjunto de itens existentes e de um conjunto de regras que podem ser utilizadas para selecionar alguns itens desse conjunto, há diferentes maneiras de escolher esses elementos e criar outros conjuntos menores (subconjuntos).
- Considerando que a cada elemento está associado um esforço ou custo, os subconjuntos criados, também, terão um custo ou esforço, que é dado, por exemplo, pela soma dos custos de seus elementos.
- Um problema de otimização combinatória se resume a encontrar, dentre os possíveis conjuntos, aquele cujo custo seja o menor possível. Um exemplo clássico é o problema do caixeiro viajante, quer precisa selecionar um roteiro para percorrer um determinado número de localidades dispersas em um território. Seu desafio é justamente encontrar o caminho "ótimo", ou seja, aquele que demanda menor custo e menor esforço.

Observando o comportamento das formigas na busca pelos alimentos, percebeu-se que, inicialmente, cada formiga segue um caminho aleatório. Após algum tempo, elas tendiam a seguir um único caminho, considerado ótimo. Cada formiga utiliza uma comunicação indireta para indicar para as outras o quão bom foi o caminho que ela escolheu Para isso elas espalham uma substância chamada "feromônio". Em um experimento, colocando-se um ninho de formigas em um aquário com uma fonte de alimentos na outra ponta. Para chegar até esse alimento foram criados dois caminhos, sendo um maior que o outro. Como as formigas que escolheram o menor caminho faziam o percurso mais rapidamente que as outras, elas acabavam depositando uma maior quantidade de feromônio nesse caminho em relação ao outro em um mesmo

instante de tempo. Logo, em um determinado momento a intensidade do feromônio no caminho mais curto estará tão alta que quase todas as formigas seguirão por ele (FRANÇA, 2006).

DORIGO & GAMBARDELLA (1997) perceberam que as formigas resolviam um problema muito similar ao do caixeiro viajante e, inspirado nesse comportamento, apresentaram propostas para modelá-lo computacionalmente e verificar como se comportava em algumas instâncias conhecidas do problema.

Em conjunto com os algoritmos genéticos e outras técnicas, os sistemas fórmicos são classificados como "meta-heurísticas". Meta-heurísticas são paradigmas de desenvolvimento de algoritmos heurísticos para tratar problemas considerados complexos, tais como: roteirização de veículos, mapeamento genético, construção de quadros de horários para professores em escolas, corte de materiais, empacotamento de produtos, escalonamento de recursos (trabalho humano e máquinas), dentre outros.

6.3.8 Mineração de Dados, OLAP e Data Warehousing

Esses assuntos serão tratados no próximo tópico e também no capítulo 7.

6.4 Sistemas Dedicados à Descoberta de Conhecimento

6.4.1 Visão Geral

Os processos de descoberta de conhecimento surgiram no âmbito da área de inteligência artificial, embora tenham recebido modernamente a contribuição da área de Banco de Dados com as estratégias de *data warehousing* e modelagem multidimensional (*OLAP*). Possuem também contribuição da Estatística. De acordo com TURBAN (2003) a inteligência artificial é uma ramificação da ciência computacional que lida com modos de representação do conhecimento por meio de símbolos e números e usando a heurística (normas práticas) e não somente de algoritmos. Segundo esse autor:

"A IA é frequentemente associada ao conceito de conhecimento. O computador não pode ter experiências nem estudar e aprender como isso é possível para a mente humana, mas pode utilizar o conhecimento[53] a ele fornecido por especialistas humanos. Esse conhecimento consiste de fatos, conceitos, teorias, métodos heurísticos, procedimentos e relacionamentos. (p. 402)."

É importante considerar que pessoas e organizações acumulam, ao longo do tempo, bases de dados ou documentos, muitas vezes perdendo a visão do que nela está inserido e da potencialidade de geração de conhecimento que essas bases possuem quando consideradas para um tratamento sinérgico por recursos de software de descoberta de conhecimento. Sistemas de Descoberta de Conhecimento são *softwares* que usam o conhecimento representado explicitamente para resolver problemas e apoiar decisões, partindo do pressuposto de que bases de dados ou de documentos embutem um conhecimento implícito, latente, escondido em suas estruturas e conteúdos. Grande parte desse conhecimento escondido é novo e relevante, pois é fruto de correlações entre os objetos de informações considerados em seu conjunto e não isoladamente. Através de uma visão sinótica, integrada e multidimensional dos elementos inseridos nessas bases, descobre-se esse conhecimento latente. Portanto, são sistemas que manipulam a informação de forma inteligente, gerando conhecimento novo.

Por descoberta de conhecimento entendemos o processo de utilização de diversas tecnologias integradas que nos permitem armazenar, tratar e recuperar informações relevantes que orientam o comportamento das pessoas, o posicionamento estratégico e a tomada de decisão nas organizações. No ambiente acadêmico, esse processo se consolidou sob a sigla KDD – *Knowledge Discovery in Database* – ou Descoberta de Conhecimento em Bases de Dados, comercialmente chamado de Mineração de Dados. A expressão mineração de dados tem origem na Estatística, na qual, nos anos 60, os profissionais e estudiosos da área se referiam como pescaria ao processo de exploração descontrolada dos dados – o que já não acontece atualmente, uma vez que os processos são bem mais definidos (PRADO, 1997). WIVES (2002) lembra que o termo mineração de dados é utilizado muitas vezes como sinônimo de

---

[53] Conhecimento aqui tem o sentido de conhecimento explicito, codificado. É a informação organizada e analisada para torná-la compreensível e aplicável à solução dos problemas e tomada de decisões, e para incoroporar procedimentos, idéias e experiência humana. TURBAN (2003, p. 402).

todo o processo de descoberta de conhecimento, mas na verdade ele é apenas o núcleo do processo, respondendo pela extração de padrões e modelos nos dados ou documentos analisados e representando cerca de 20% do processo total de descoberta de conhecimento, o qual possui outras etapas. Considerando a grande disponibilidade de dados acumulados nas organizações, os primeiros processos informatizados visando a descoberta de conhecimento foram aplicados sobre dados em geral mais estruturados (daí a expressão KDD). Posteriormente é que tais processos passaram a considerar os documentos menos estruturados, como os textos, originando a área de Descoberta de Conhecimento em Textos ou Mineração de Textos ou ainda KDT – *Knowledge Discovery in Text*, a ser detalhada no próximo capitulo.

De acordo com FAYYAD, PIATETSKY & SMYTH (1996) [54], extrair conhecimento de bases de dados é o processo de identificar padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados, sendo que todo processo é orientado em função de seu domínio de aplicação e das bases de dados ou documentos relativos a esse domínio. REZENDE (2003, p. 310) lembra que essa definição merece um detalhamento do seu conteúdo, e que podemos então esclarecer alguns dos conceitos que nela aparece da seguinte forma:

- Padrões – indica alguma abstração de um subconjunto dos dados em alguma linguagem descritiva de conceitos.
- Processo – a descoberta de conhecimento em bases de dados possui etapas específicas a serem seguidas.
- Válidos – os padrões descobertos devem possuir algum grau de certeza, ou seja, devem satisfazer funções ou limiares que garantam que os exemplos cobertos e os casos relacionados ao padrão encontrado sejam aceitáveis.
- Novos – um padrão encontrado deve fornecer novas informações sobre os dados. O grau de novidade serve para determinar quão novo ou inédito é um padrão. Pode ser medido por meio de comparações entre as mudanças ocorridas nos dados ou no conhecimento anterior.

---

[54] FAYYAD, U., PIATETSKY-SHAPIRO, G. & SMYTH, P. From data mining to knowledge discovery: na overview. In *Advances in Knowledge Discovery & Data Mining*. 1996, p. 1-34. Citação apresentada em REZENDE (2003, p.309-310).

- Úteis – os padrões descobertos devem ser incorporados para serem utilizados.
- Compreensíveis – um dos objetos de se realizar a mineração de dados é encontrar padrões descritos em alguma linguagem que pode ser compreendida pelos usuários permitindo uma análise mais profunda dos dados.

6.4.2 O Processo de Descoberta de Conhecimento em Bases de Dados

Analisando a obra de WESTPHAL & BLAXTON (1998), podemos dizer que a Descoberta de Conhecimento, Mineração de Dados e mesmo a Mineração de Textos são expressões relativas a um processo amplo que inclui o uso de uma ou mais ferramentas (*softwares*) "inteligentes" (que se servem de um ou vários recursos da chamada Inteligência Artificial, da Estatística e da área de Banco de Dados) e que viabiliza a busca de dados em bases de uma empresa, procurando descobrir informações desconhecidas pelos tomadores de decisão. São ferramentas que atuam na descoberta de informações articuladas que expressam "conhecimento" obtido pelo cotejamento e organização dos dados (base de dados), centralizando-se na descoberta automática de novos fatos e relações nos dados[55].

O processo de descoberta de conhecimento em base de dados possui quatro passos:

a) Identificação do Problema e Definição do Domínio da Aplicação: É o estudo do domínio da aplicação e a definição de metas a serem alcançadas no processo de mineração (REZENDE, 2003). Especialistas conhecedores do domínio a ser alvo da descoberta do conhecimento precisam participar dessa fase para apoiar o processo. Definem-se, além das metas, os critérios de desempenho, os aspetos de compreensibilidade do conhecimento extraído e as questões relativas à precisão e usabilidade desse conhecimento. É também importante definir quais as fontes de dados a ser usadas nas etapas seguintes.

b) Preparação ou pré-processamento: como o próprio nome sugere é nesse passo que os dados são preparados antes de serem apresentados a um agente inteligente de software.

[55] Ver também, para uma análise mais acurada sobre *Data Mining* a obra de JIAWEI, H. e KAMBER, M. *Data Mining: concepts and Tecniques*. San Francisco, USA. Simon Frise University, 2001.

Os dados importantes são selecionados, purificados (eliminação de inconsistências) e pré-processados com definição de quais dados são importantes e o que fazer com eles.

c) Extração de Padrões ou Mineração (*Mining*) é o processamento dos dados com a mineração propriamente dita. Temos diferentes tipos de descobertas possíveis a serem realizadas aqui. Os dados são aqui transformados em informações importantes e o foco e conseguir gerar conhecimento novo e útil. Esse passo é o núcleo do processo e será detalhado no tópico 6.4.3

d) Análise ou Pós-Processamento: O resultado da mineração é avaliado, visando determinar a importância dos fatos gerados e se algum conhecimento adicional foi descoberto. De acordo com WIVES (2002), o pós-processamento implica também em aplicação de filtros de estruturação e ordenação para que o conhecimento possa ser apresentado ao usuário da forma mais simples e compreensível possível.

Para GOEBEL & GRUENWALD (1999) o processo (computacional) de descoberta de conhecimento possui as seguintes etapas:

- entendimento do domínio da aplicação e definição dos objetivos do processo de descoberta;
- aquisição ou seleção do conjunto dos dados;
- integração e verificação do conjunto;
- limpeza dos dados (pré-processamento e transformação);
- desenvolvimento de um modelo inicial ou construção de hipóteses iniciais;
- escolha e aplicação de métodos (tarefas) de mineração dos dados;
- visualização e interpretação dos resultados;
- teste e validação das hipóteses (pode-se refazer parte do processo;
- uso e manutenção do conhecimento descoberto (tomada de decisão no domínio da aplicação).

Detalhando as questões acima discutidas, uma visão resumida da seqüência dos passos do processo de descoberta de conhecimento em bases de dados pode ser apresentada na figura 26 a seguir:

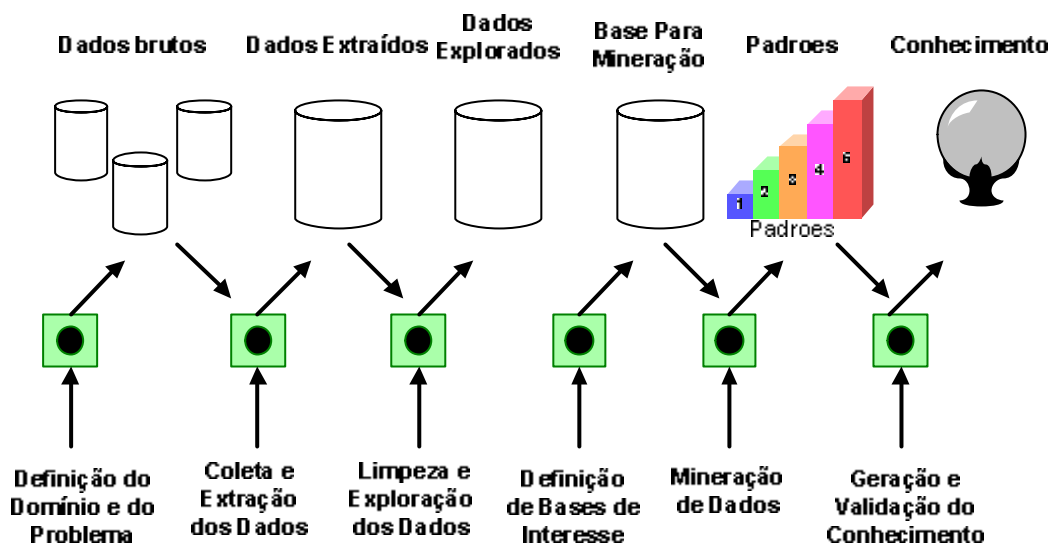



Figura 26: O processo de Descoberta de Conhecimento em Bases de Dados
Fonte: Adaptado de PRADO (1997, p. 12)

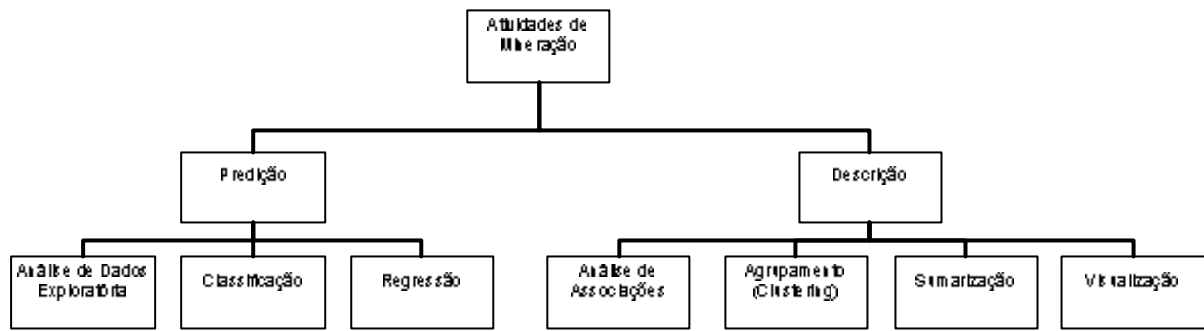
6.4.3 Tarefas de Mineração de Dados

De acordo com GOEBEL & GRUENWALD (1999), o aspecto central dos processos de mineração de dados é a extração de padrões dos dados. Para isso diferentes métodos ou tarefas podem ser usados para diferentes objetivos. Inicialmente, é necessário definir qual tarefa de mineração de dados será usada, o que é feito de acordo com o objetivo estabelecido para o processo. Esses autores relacionam oito tarefas de mineração de dados:

- Processamento dos Dados: tarefa que tem a finalidade com o intuito de selecionar, filtrar, agregar, exemplificar, limpar e transformar os dados.
- Análise de Dados Exploratória: é a exploração interativa de dados sem modelos ou hipóteses antecipados, com objetivo de se identificar padrões de comportamento, traços e tendências embutidos nos conjuntos de dados.
- Classificação: tem por objetivo determinar a classe a que um certo dado pertence. A classificação é uma função de aprendizado que organiza os dados em categorias. Ele visa enquadrar em uma classe qualquer exemplo novo que está sendo examinado.

- Regressão: são atividades que avaliam a dependência de valores de certos atributos em relação a outros do mesmo item, construindo um modelo com capacidade de prever os valores de novos registros. A tarefa de regressão é similar à de classificação. A diferença central reside no fato de que o atributo a ser predito é continuo em vez de discreto.
- Análise de Associações (ou *Links*): Os dados ou documentos podem possuir elos (*links*) entre eles, indicando sua similaridade. Essa tarefa tem por objetivo identificar relacionamentos entre atributos e itens, com a finalidade de perceber se a presença de algum padrão redunda na presença de outro.
- Agrupamento (*Clustering*): dividem o conjunto de itens em exame em subgrupos com características similares. Implica na segmentação da informação disponível em conjuntos definidos e homogêneos com base em atributos específicos.
- Sumarização: é a geração de descrições resumidas para um subconjunto de dados – resumos automáticos.
- Modelo de Visualização: objetiva tornar o conhecimento mais palatável e compreensível para o usuário através de gráficos, animações, representações em mapas, desenhos tridimensionais, etc.

Esse processo pode ter caráter preditivo ou descritivo, de acordo com REZENDE (2003). A predição implica generalizar exemplos ou experiências passadas com respostas conhecidas em uma linguagem, capaz de reconhecer a classe de um novo exemplo. As atividades de predição englobam o uso dos atributos de um conjunto de dados para prever o valor futuro de uma determinada variável. Por sua vez, a descrição é um conjunto de atividades que visam descobrir padrões que possam ser interpretados pelas pessoas que descrevem os dados antes de realizar a previsão. A descrição é a identificação dos comportamentos específicos do conjunto dos dados, sendo estes dados não possuem uma classe especificada. A estrutura abaixo, adaptada de REZENDE (2003, p. 318) e GOEBEL & GRUENWALD (1999, p. 22) aponta esses processos, atividades ou tarefas de mineração de dados.

Atividades de Mineração
Predição
Descrição
Análise de Dados Exploratória
Classificação
Regressão
Análise de Associações
Agrupamento (Clustering)
Sumarização
Visualização

- Na etapa inicial, de entendimento do domínio da aplicação e definição do objetivo do processo de descoberta, percebemos que na busca proativa pode não haver uma definição precisa do objetivo do processo, ou seja: há um objetivo ou problema, mas a solução não é possível de ser prevista.
- No que tange ao desenvolvimento de modelos iniciais ou construção de hipóteses iniciais, na abordagem proativa não é possível fazer isso, uma vez que o usuário não sabe ou não quer elaborar hipóteses para a solução de um problema.

Baseando-se em KUHLTHAU (1991), podemos dizer que em um processo exploratório de busca proativa, o usuário começa tentando encontrar alguma forma de conhecimento mais geral, indo posteriormente para a busca de conhecimentos mais restritos, com informações mais direcionadas ou específicas. Ele então tenta reconhecer, identificar, formular, e reunir e complementar diferentes elementos que constituem o conhecimento de seu interesse.

6.4.5 Abordagens Tecnológicas Utilizadas em Descoberta de Conhecimento em Bases de Dados

O processo de descoberta de conhecimento em bases de dados engloba várias abordagens tecnológicas, tais como as descritas anteriormente, especialmente no tópico 6.3 dessa tese (aprendizagem de máquina, algoritmos genéticos, lógica difusa, redes neurais, árvores de decisão, redes *bayesianas*, dentre outros), bem como as estratégias de *data warehousing* e modelagem multidimensional, descritas no capitulo 7 desse trabalho.

6.5 Processamento da Linguagem Natural – PLN

Esse é um campo que interessa, sobretudo, aos sistemas de recuperação de informação, ou seja: sistemas que lidam com documentos textuais. O Processamento da Linguagem Natural – PLN é uma subárea da Inteligência Artificial – IA, resultado da interseção da lingüística moderna com as tecnologias desenvolvidas no campo da IA. De acordo com RUSSEL & NORVIG (2004), a compreensão da linguagem exige a compreensão do assunto e do contexto, não apenas a compreensão da estrutura das

frases. O PLN se baseia em idéias de lingüística e filosofia, além de técnicas de representação do conhecimento lógico e probabilístico de raciocínio. Essa área exige uma investigação do comportamento humano real, o que amplia o desafio. O PLN é um conjunto de ações que engloba o estudo, planejamento e desenvolvimento de dispositivos computacionais relativos ao uso da linguagem, ou seja: a análise e projeto de agentes computacionais que usam a linguagem natural para adquirir informação.

A idéia de linguagem está vinculada à necessidade de comunicação, que segundo RUSSEL & NORVIG (2004) pode ser entendida como:

> "*a troca intencional de informações provocada pela produção e percepção de sinais extraídos de um sistema compartilhado de sinais convencionais. O que distingue o ser humano dos outros animais é o complexo sistema de mensagens estruturadas conhecido como linguagem e que nos permite comunicar a maior parte daquilo que sabemos sobre o mundo* (p. 765)".

Ainda de acordo com esses autores:

> "*Os agentes*[56] *enviam sinais uns aos outros para atingir certos propósitos: informar, advertir, pedir ajuda, compartilhar conhecimento ou prometer algo. O envio de um sinal dessa maneira é chamado ato de fala. Em última instância, todos os atos da fala são tentativas de fazer outro agente acreditar em algo ou realizar algo. A linguagem consiste em sinais convencionais que transmitem significado. (...) Os seres humanos parecem ser os únicos animais a utilizar a gramática para produzir uma variedade ilimitada de mensagens estruturadas. A comunicação envolve três etapas seguidas pelo falante: a intenção de transmitir uma idéia, a geração mental de palavras e sua síntese física. Em seguida, o ouvinte segue quatro etapas: percepção, análise, eliminação de ambigüidade*[57] *e incorporação do significado. Toda utilização de linguagem está situada, no sentido de que o significado de uma expressão vocal pode depender da situação em que ela é produzida* (p.799)".

De acordo com LANCASTER (2004) considera-se a linguagem natural como sinônimo de discurso comum, ou seja: a linguagem usada normalmente na fala e na

---

[56] Conforme já citado, um agente é algo que percebe e age em um ambiente. Pode ser identificado com o ser humano, em um contexto social real, ou um dispositivo computacional que "tenta imitar" o comportamento humano. É um elemento capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por meio de atuadores. A inteligência artificial é, para tais autores, uma ciência de projeto de agentes inteligentes, ou seja: agentes bem sucedidos para a função que se propõe e que envolvem a intenção de simular ou atuar de forma análoga aos processos cognitivos e racionais humanos

[57] Palavras e expressões podem assumir diferentes significados e posições no discurso. A multiplicidade conceitual ou categorial que uma palavra pode assumir (ambigüidade léxica) ou a multiplicidade estrutural na representação sintática de uma frase (ambigüidade estrutural). A superação da ambigüidade exige um conhecimento do contexto, do ambiente e das relações de um discurso ou fala, bem como dos textos escritos.

escrita. Uma linguagem natural, como a língua portuguesa, é usada por um grupo humano específico e normalmente concede uma liberdade muito grande de uso quando comparada a uma linguagem inventada ou formal. Essa liberdade de uso significa que é possível representar um conjunto muito grande de mensagens e idéias. As linguagens consistem de dois elementos básicos: um vocabulário e uma estrutura gramatical[58] ou sintaxe.

De acordo com RUSSEL & NORVIG (2004, p. 767), uma gramática é um conjunto finito de regras que especificam uma linguagem (p.767). Conforme CANÇADO (2005) gramática é o conhecimento da língua, ou o sistema de regras e princípios que governam o uso dos signos da língua. De acordo com essa autora:

> "*a lingüística assume que o falante de qualquer língua possui diferentes tipos de conhecimento em sua gramática: o vocabulário adquirido, como pronunciar as palavras, como construir as palavras, como construir as sentenças e como entender o significado das palavras e das sentenças. Refletindo essa divisão, a descrição lingüística tem diferentes níveis de análise: o léxico, que é o conjunto de palavras de uma língua; a fonologia, que é o estudo dos sons de uma língua e de como esses sons se combinam para formar as palavras; a morfologia, que é estudo das construções das palavras; a sintaxe, que é o estudo de como as palavras podem ser combinadas em sentenças; e a semântica, que é o estudo do significado das palavras. Ao conhecimento da língua, associa-se o uso da língua, ou seja, o emprego da gramática dessa língua nas diferentes situações, numa área denominada pragmática, que estuda a maneira pela qual a gramática, como um todo, pode ser usada em situações comunicativas concretas*" (p. 15-16).

Nessa citação, merece um comentário o conceito de sentença, não só pela sua importância lingüística, mas também como elemento a ser tratado pelo processamento computacional. De acordo com PIRES DE OLIVEIRA (2001, p. 99) uma sentença (S) pode ser definida sintaticamente pela presença de um verbo principal conjugado e, semanticamente, pela expressão de um pensamento completo. O processamento da linguagem natural, portanto, implica em que processos computacionais devem considerar todos esses aspectos lingüísticos acima apontados. Um sistema clássico de PLN pode envolver:

---

[58] De acordo com RUSSEL & NORVIG (2004), uma gramática é um conjunto finito de regras que especificam uma linguagem (p.767). Conforme CANÇADO (2005) gramática é o conhecimento da língua, ou o sistema de regras e princípios que governam o uso dos signos da língua.

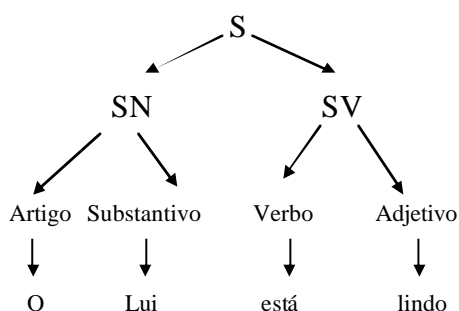
- Uso de dicionários – que considere palavras e suas estruturas – bases e terminações (raiz, prefixos e sufixos)
- Uso de gramáticas
- Presença de analisador léxico-morfológico
- Presença de analisador sintático
- Presença de analisador semântico – grande desafio
- Presença de analisador pragmático
- Presença de analisador do discurso

SALTON (1983)[59] propõe modelos matemáticos para especificar os processos lingüísticos e métodos computacionais para descrever a interação entre estes modelos. Esse autor, em coerência com a visão lingüística supracitada, define diferentes níveis de processamento da linguagem natural, ou do conhecimento embutido na linguagem, quais sejam o fonético, o morfológico, o léxico, o sintático, o semântico e o pragmático. Esses níveis são comentados a seguir, conforme as contribuições dadas por ele, incorporando contribuições de outros autores na análise.

- Nível Fonético – lida com os sons e com a compreensão da fala.
- Nível Morfológico – analisa a forma e inflexão das palavras. Procura lidar com o processamento de partes reconhecíveis das palavras, realizando análise e/ou retiradas de prefixos, sufixos e raízes (*steams*) das mesmas.
- Nível Léxico – tratamento em nível de itens lexicais. Lida com procedimentos sobre as palavras completas. Trata do processamento de dicionários, substituições de palavras por classes de tesauros, por exemplo. Lida com informações gramaticais, como natureza e flexão. Na análise sintática, identifica características lingüísticas, como substantivos, adjetivos, verbos, preposições, etc.
- Nível Sintático – lida com as estruturas sintáticas. Identifica a estrutura da frase por meio das regras de sintaxe da língua. Analisa a disposição dos itens lexicais em uma sentença, e das relações existentes entre estes. Promove a decomposição de um período, e de cada oração, em seus elementos (termos ou sintagmas). Um determinado termo corresponde a um elemento de informação (contendo itens lexicais ou palavras) e é tratado como unidade funcional da oração, participando de sua estrutura como um de seus constituintes, chamados sintagmas (unidades sintáticas), de acordo com suas funções. Em geral, nesse nível é são verificadas

[59] Abordagem detalhada em SALTON (1983, Cap. 7).

regras de formação impostas pela linguagem, utilizando-se das informações léxico-morfológicas disponíveis, tais como: concordância nominal e verbal, regência nominal e verbal e posicionamento dos termos na frase. Para RUSSEL & NORVIG (2004) a análise sintática é o processo de construir uma árvore de análise para uma cadeia de entrada, sendo que os nós interiores da árvore de análise representam os sintagmas e os nós de folhas representam as palavras. Essa idéia é representada árvore estrutural de uma sentença S (abaixo). Os componentes envolvidos na análise sintática são:

a. uma representação gramatical, onde estão declarados os fatos sintáticos da linguagem, as estruturas de composição dos sintagmas;
b. um procedimento de análise, responsável por verificar se a frase de entrada está de acordo com estas regras gramaticais, gerando uma estrutura hierárquica que representa a estrutura da frase analisada.

Na representação acima, S é uma sentença – "*O Lui está lindo*". SN é a indicação de um sintagma nominal e SV um sintagma verbal. Conforme RUSSEL e NORVIG (2004), a maior parte dos formalismos gramaticais se baseia na idéia das estruturas acima, que podem ser chamadas de estruturas sintagmáticas, em que a cadeia é composta por subcadeias, chamadas de sintagmas, que são elementos semânticos naturais a partir dos quais o significado de uma expressão vocal pode ser construído. KURAMOTO (2002) avaliou a limitação do uso das palavras como meio de acesso à informação pelos sistemas automatizados de recuperação de informação, afirmando que apesar de alguns deles terem alcançado relativo sucesso na melhoria da precisão de resultados de uma busca, a meta principal da recuperação de informação, que é a obtenção de todos os documentos pertinentes a uma consulta, não foi atingida.

Por isso esse autor advoga a inadequação do uso das palavras nesses modelos, propondo em seu lugar, um outro tipo de unidade de informação os sintagmas nominais, conceito da lingüística tratado no processamento da linguagem natural.

- Nível Semântico – nível em que os elementos da estrutura sintática são transportados para uma estrutura de representação de conhecimento. A preocupação principal é o significado. De acordo com RUSSEL & NORVIG (2004) a interpretação semântica é o processo de extrair o significado de uma expressão vocal (ou escrita) como uma expressão em alguma linguagem de representação. O tratamento semântico é o processo de mapeamento das sentenças para uma representação de seu significado.
-
- Nível Pragmático – lida com o contexto, ou seja, com informações adicionais sobre o ambiente social em que documentos textuais, procurando fazer a efetiva identificação do que realmente se quis ou se quer dizer. Busca-se entender o real sentido do discurso, no qual o significado capturado na frase é integrado ao contexto. De acordo com RUSSEL & NORVIG (2004),

> "*na interpretação pragmática leva-se em conta o fato de que as mesmas palavras podem ter significados diferentes em situações diferentes. Enquanto a interpretação sintática é uma função de um único argumento, a cadeia, a interpretação pragmática é uma função de expressão vocal e do contexto ou situação em que é articulada* (p. 769)".

O uso do processamento de linguagem natural - PLN pode ser dividido em dois grupos, de acordo com ALLEN (1994): uso baseado em texto e uso baseado em diálogo. O uso baseado em texto, implica na busca de documentos num banco de dados de textos, extração de informação de textos, tradução de documentos, resumos de textos e compreensão de textos. Os recursos usados para operar tais aplicações são conhecidos como recursos ou técnicas de PLN, sendo que o foco de interesse dessa Tese é o processamento de documentos textuais.

7 TECNOLOGIA DA INFORMAÇÃO E GESTÃO DO CONHECIMENTO

7.1 Considerações Preliminares

No ambiente organizacional do atual cenário mundial, as pressões por competitividade e busca de um posicionamento estratégico adequado provoca uma demanda por capacidades de lidar com a informação que é respondida com uso cada vez mais intenso de tecnologias baseadas em recursos computacionais. Na medida em que o conhecimento está ficando cada vez mais importante na geração de riqueza (STEWART, 1998), precisamos olhar para as empresas enquanto gestoras de conhecimento. Muitos gestores, inseridos num ambiente saturado de dados, presenciando um ritmo acelerado para produzir, controlar e "explorar rapidamente o conhecimento", percebem que necessitam de mais informação sobre sua realidade organizacional e o contexto ambiental e mercadológico em que se encontram. Essa necessidade sempre existiu, mas agora se intensifica, considerando fatores, tais como:

- a complexidade tecnológica para operação de muitas organizações modernas;
- o quanto o conhecimento especializado de muitas pessoas é decisivo para o sucesso da missão da organização;
- o aumento da concorrência no mundo inteiro;
- a facilidade de se sobrepor ao mercado do concorrente, até mesmo se ele estiver atuando do outro lado do mundo;
- o tempo de durabilidade do produto que está se tornando mais e mais reduzido.

Fazer uso tempestivo e eficaz, tanto de bases de dados relacionais dos sistemas transacionais e gerenciais, bem como dos sistemas que lidam com documentos é portanto um desafio.A implementação das propostas, tanto dos paradigmas dos sistemas baseados em bancos de dados como dos sistemas baseados em recuperação da informação são implementadas em determinadas plataformas e ambientes tecnológicos. Hoje, e cada vez mais, tais sistemas são baseados em recursos computacionais. A tecnologia da informação inclui um conjunto de elementos, de caráter computacional e eletrônico, que viabilizam a captura, armazenagem, organização e tratamento dos dados de uma organização de forma a transformá-los em informações úteis para uma organização. São inseridos aqui todos os componentes de hardware, categorias

diferentes de software e metodologias de tratamento de dados e geração de informação útil para as organizações, sejam elas gerenciais ou operacionais. A tecnologia da informação inclui também todos os processos de recuperação e disponibilização de informações para o seu destinatário final. Obviamente possuímos diferentes tipos de computadores, categorias de software, metodologias, recursos de telecomunicações, etc. que se inserem na categoria de tecnologias da informação. Devemos lembrar sempre suas muitas e diferentes configurações em cada organização, dependendo de seus interesses e seu fim último: gerar informação útil.

Os fluxos de informação são hoje altamente dependentes da tecnologia da informação. A Tecnologia da Informação vem impulsionando fortemente o desenvolvimento da Ciência da Informação, ampliando suas possibilidades e valorizando suas conquistas, o que vale obviamente para a recuperação da informação e para a gestão do conhecimento Para SARACEVIC (1992) há uma conexão "inexorável" entre ambas. SALTON (1985) indica que construções importantes da Ciência da Informação possuem aplicações práticas, especialmente vinculadas a sistemas de informação, tais como modelos *booleanos* na recuperação da informação, modelos de realimentação de relevância para recuperação, redes de citação, análise lingüística para indexação automática, dentre outros. Além disso, os campos de estudo dos tesauros, da terminologia e da classificação apóiam sistemas e atividades que se ocupam da organização do conhecimento em suas diferentes manifestações como, por exemplo, os sistemas de inteligência artificial e hipertextos. A mineração de textos terá também uma forte articulação com o desenvolvimento da Ciência da Informação, servindo-se de suas construções e ampliando suas propostas.

Para DRUCKER (1988), as empresas não possuem outra escolha a não ser lastrear toda sua atividade em informação. A informação adequadamente identificada para uma necessidade específica, produzida a partir da captura de dados em fontes precisamente mapeadas, metodologicamente tratada para atender propósitos aos quais se destina e disponibilizada de forma tempestiva e no contexto pertinente de interação almejado pelo usuário é um elemento importante de vantagem competitiva nas organizações. Para fazer com que a informação possa ser efetivamente utilizada nas organizações, assegurando um processo decisório adequado, faz-se mister o uso de

tecnologias da informação. Obviamente que a obtenção de informações não é uma tarefa fácil, tem custos não desprezíveis.

A tecnologia é uma variável das mais importantes a proporcionar conjuntos relevantes de transformações nas empresas contemporâneas. Há uma aceleração do processo de geração de inovações as quais, por sua vez, tem raízes no processo de geração de informação e conhecimento e a tecnologia pode ser considerada um lastro desse processo. Tecnologia da informação é também um instrumento poderoso na viabilização de novas estratégias empresariais. Conforme afirmam NADLER et alli (1994):

> "*a tecnologia da informação começou a revolucionar o projeto organizacional ao proporcionar alternativas à hierarquia como meio principal de coordenação. Sistemas de informação, arquiteturas comuns, bancos de dados compartilhados, ferramentas de apoio às decisões e sistemas especialistas facilitam a coordenação do comportamento sem controle por meio da hierarquia, permitindo com isso a criação de unidades autônomas ligadas pela informação. Também permitem um acoplamento menos rígido (em oposição ao acoplamento rígido) sem os riscos de perda de coordenação e controle. A combinação do grande potencial de tecnologia da informação com as grandes exigências do ambiente competitivo levou a inovações no projeto organizacional* (NADLER et alli, 1994)"

A tecnologia da informação é um campo interdisciplinar que, apoiado na Ciência da Computação, na Engenharia de Software, Engenharia de Computação, na Administração e na própria Ciência da Informação, busca construir uma base científica para assuntos diversos, tais como a teoria das decisões, o estudo e solução de problemas com algoritmos, a elaboração de programas e desenvolvimento de sistemas, o desenvolvimento de dispositivos e máquinas computacionais eletrônicas, o estabelecimentos de redes digitais, dentre outros. O processamento automatizado de dados a serem convertidos em informações, especialmente pela captura de algum nível de inteligência por algoritmos traduzíveis em programas de computador, fez os aspectos computacionais merecerem papel de destaque nas últimas décadas.

Conforme BROOKSHEAR (2000), os algoritmos têm um papel relevante para a computação e a tecnologia da informação. Outrora estudados pela matemática, seu estudo ganhou maior robustez com o advento dos computadores, visando encontrar um conjunto de comandos que fosse capaz de descrever, de maneira unificada, a resolução

de qualquer problema de determinado tipo. Em seguida, esses algoritmos precisam ser representados computacionalmente para que possa ser executado por um software em alguma máquina. Conforme esse autor:

> *"uma vez encontrado um algoritmo para uma determinada tarefa, sua execução já não dependerá do conhecimento dos princípios nos quais se baseia, restringindo-se apenas a seguir instruções estabelecidas. Em outras palavras, o algoritmo constitui uma codificação do raciocínio necessário á resolução do problema. É por meio dessa capacidade de captar e transferir inteligência mediante os algoritmos que são construídas máquinas com comportamento inteligente. Por conseguinte, o nível de inteligência demonstrado pelas máquinas fica limitado pela inteligência que um algoritmo é capaz de transportar. Somente quando for possível obter um algoritmo que possa controlar a operação de uma tarefa será viável construir alguma máquina capaz de executá-la. Por outro lado, se não houver algoritmo capaz de executar a tarefa, então sua capacidade excederá as capacidades da máquina. (...) Uma vez descoberto um algoritmo para solucionar o problema, o passo seguinte consiste em representa-lo de forma apropriada para que seja transmitido a alguma máquina, ou para que seja lido por outros seres humanos. Isso significa que se torna necessário transformar o algoritmo conceitual em um conjunto facilmente compreensível de comandos que representem, sem ambigüidade, essas instruções"* (p16-17).

Nesse contexto, estudos fundamentados em um conhecimento lingüístico e gramatical conduziram a uma grande diversidade de esquemas para a representação de algoritmos, as chamadas linguagens de programação, baseadas em visões variadas (paradigmas)[60] do processo de programação. Esses aspectos demandaram recursos importantes para sua implementação, exigindo adequados projetos para construção de máquinas (hoje tratados especialmente pela chamada Engenharia da Computação, focada no projeto e desenvolvimento de equipamentos computacionais) e demandando estratégias adequadas para desenvolvimento de grandes sistemas automatizados, o que foi respondido pelo advento da chamada Engenharia de Software.

Destarte, assumindo uma perspectiva organizacional, podemos dizer que a tecnologia da informação inclui um conjunto de elementos que viabilizam a captura, armazenagem, organização e tratamento dos dados de uma organização de forma a transformá-los em informações úteis para uma organização. Inclui todas as estratégias tecnológicas de geração, armazenagem recuperação e entrega de informações (ou dados, dependendo do ponto de vista) para o seu destinatário ou interessado final. É o resultado da convergência de três vetores que se encontram em três planos distintos:

[60] Paradigmas de linguagens: imperativas, declarativas, funcionais e orientadas a objeto, por exemplo.

conceitual, lógico e físico. Cada um desses planos contém seus elementos específicos e que podem ser caracterizados conforme a seguir.

- Plano Conceitual. Elementos Conceituais e Metodológicos: abordagens e critérios para desenho da arquitetura lógica de sistemas de informação e desenvolvimento de software. Inclui todas as ferramentas de modelagem de dados, construção de sistemas e desenho de seu projeto conceitual e lógico. Essas abordagens se apóiam em algum dos paradigmas descritos nos capítulos anteriores. Como, por exemplo, dentre outras, temos as diversas abordagens para lidar com a organização de dados, como o modelo relacional, as propostas para desenvolvimento de sistemas, tais como a Análise Estruturada, o Processo Unificado para a Orientação a Objetos e as abordagens de planejamento da arquitetura da informação estratégica, tais como os Fatores Chave de Sucesso - FCS's e o *Balanced Scorecard* - *BSC*. Também podemos considerar os modelos dos sistemas de recuperação da informação, descritos no capitulo 5, dentro dessa perspectiva.

- Plano Lógico. Os Softwares: todos os sistemas desenvolvidos sob forma de programação computacional articulada para atender a um determinado objetivo e com um conjunto de funções previamente definidas. Na Ciência da Computação, algoritmo é um conceito fundamental e no domínio dos computadores os algoritmos são representados através dos programas. Um software é um conjunto de programas que, por sua vez, expressam algoritmos, visando atender a uma determinada finalidade. Incluem-se aqui todos os softwares desenvolvidos por uma organização ou adquiridos de terceiros. Esses softwares podem ser aplicativos específicos para melhorar ou automatizar funções organizacionais, bem como *softwares* de suporte essencial, como os bancos de dados, navegadores da *Internet* e as linguagens de programação.

- Plano Físico. A Infra-estrutura ou Hardware: inclui todos os equipamentos e recursos físicos utilizados para o processamento, tratamento e disponibilização de dados. Computadores pessoais, servidores ou *mainframes*, cabos e redes de fibra óptica, *modens* e impressoras são bons exemplos. Com o advento e aprofundamento do uso da Internet é importante considerar, também, estruturas que no passado classicamente não se ajustavam a essa categoria agora passam a fazer parte dela cada

vez mais, como a infra-estrutura de telecomunicações caracterizando o fenômeno da convergência digital – possibilidade aberta pela interseção da Tecnologia da Informação com as tecnologias de comunicação de dinamizar não só a oferta de informação, mas facilitar sua publicação, proporcionando interatividade.

A tecnologia da informação nas organizações começou a ser utilizada nos anos 60 para melhorar o desempenho de tarefas básicas automatizando rotinas como as de contabilidade e folha de pagamento de pessoal. A partir de então, muitas aplicações para problemas de todas as ordens estão sendo alvo do esforço computacional em aplicações chamadas operacionais - ou seja - problemas do dia-a-dia das organizações. Com o tempo, os sistemas de informação passaram a trabalhar cada vez mais integrados uns aos outros, de forma que uma ação em um sistema de informação gerasse reflexos em outros sistemas e esse ambiente recebe o nome de transacional.[61] Transações são eventos vinculados nos sistemas de informação. Por exemplo, uma venda registrada no sistema de vendas, tem um reflexo no sistema de estoques. Por outro lado, temos também os sistemas voltados para suporte à decisão, com capacidade de gerar informação com maior valor agregado. A busca por informações estratégicas, mais elaboradas, capazes de propiciar uma melhor tomada de decisões por parte das organizações, foi uma primeira vertente para lidar com a tecnologia da informação sustentando uma estratégia de uso do conhecimento nas organizações.

A partir do final dos anos 70, ferramentas de software para apoio a decisão gerencial começam a se popularizar no ambiente organizacional. No que tange a softwares, em meados dos anos 80, a tecnologia dos Sistemas de Informação Executiva (os *Executive Information Systems* - EIS, softwares desenvolvidos especificamente para o processo de tomada de decisão gerencial e baseados em Fatores Chave de Sucesso) começou a produzir mudanças significativas no processo de gestão da informação estratégica nas organizações. No final dos anos 90, começam a surgir softwares baseados na visão do BSC.

---

[61] Uma classificação que divide os sistemas de informação em duas categorias, sistemas de processamento de transações e sistemas de suporte à decisão é proposta por DHAR,V. & STEIN, R. em *Seven Methods for Transforming Corporate Data into Business Intelligence*. USA, Prentice-Hall, 1997. A primeira categoria visa o registro detalhado das operações e fatos relevantes de todas as atividades de negócio, tendo por ênfase a validade e depuração das bases de dados. A segunda categoria apóia os gestores de negócio no processo de tomada de decisão, considerando uma visão de longo prazo.

Na verdade hoje, tanto o ambiente de informação transacional, quanto o gerencial mudaram significativamente. No campo dos sistemas transacionais, o final dos anos 90 assistiu ao crescimento dos chamados softwares de gestão integrada, os ERP's - *Enterprise Resources Planning*. Conforme COLANGELO FILHO (2001), não há uma definição precisa do que seja um sistema ERP, mas de acordo com ele é uma plataforma de software que permite:

- *"automatizar e integrar parcela substancial de seus processos de negócios, abrangendo finanças, controles, logística (suprimentos, fabricação e vendas) e recursos humanos;*
- *compartilhar dados e uniformizar processos de negócios;*
- *produzir e utilizar informações em tempo real (p. 17)"* .

TURBAN (2003) salienta que tais sistemas estão mais vinculados à atividade fim da empresa, e que visa:

> *"coordenar todos os recursos materiais, de produção e de economia global existentes dentro de uma empresa, geralmente vinculando todas as áreas funcionais que contribuem de alguma maneira para a produção de determinado produto* (p. 12)."

O conceito de ERP's está, portanto, fortemente vinculado à idéia de integração da empresa como um todo em uma única plataforma de software, com módulos *"plug in"* [62] dedicados a cada função, como compras, vendas e produção. O ERP demanda e promove a criação de uma grande base de dados corporativa nas organizações. Conforme afirma CORRÊA, et alli (1999):

> "A *tendência parece claramente indicar que as estruturas dos ERP's serão utilizadas pelas empresas como as fundações (a grande base de dados corporativa para apoio à tomada de decisão, principalmente operacional) dos sistemas de informação das empresas* (CORRÊA, et alli, 1999:344)".

Entretanto, os ERP's enfrentaram (e ainda enfrentam) muitos problemas de implantação e precisam ainda de um esforço estratégico mais denso para consolidar seu verdadeiro papel nas organizações. Focam mais aspectos da operação do que da gestão organizacional propriamente dita.

---

[62] Em geral, adquiridos sob a forma de pacotes comerciais de software.

A utilização de estratégias tecnológicas que sustentem a visão de valorização e melhor aplicação do conhecimento organizacional ganha bom impulso com o uso de tecnologias da informação, com o uso de bancos de dados, redes de computadores e metodologias de tratamento da informação, como o *Balanced Scorecard*, os Fatores Chaves de Sucesso. Da mesma forma, a construção de "*intranets*" com a mesma lógica de liberdade da rede mundial de computadores pode estimular a geração e disponibilização do conhecimento em organizações que estimulem uma visão de "organização em rede de pessoas e comunidades internas" Pessoas, áreas, atividades de projetos podem interagir entre si compartilhando suas experiências, em "sites" internos e em grupos informais. É importante, entretanto, que a estratégia possua um nível de articulação mínimo, ou seja, uma estrutura, um *framework* com intencionalidade organizacional que a permita efetivamente gerir, organizar e disseminar o conhecimento. Podemos então considerar a taxonomia como um elemento chave e que tem um componente político e ideológico na medida em que faz opções. Esse elemento político-ideológico está ligado à natureza da organização, às suas intenções estratégicas. Portanto, às suas escolhas, planos e padrões de atuação no ambiente. Por exemplo, a taxonomia do conhecimento poderia se orientar por funções organizacionais, tais como produção e marketing ou poderia se orientar por processos, tais como manufatura e comercialização. Poderia ainda ser organizada por assunto, tais como produtos, pessoas, recursos financeiros.

A perspectiva do estudo da gestão do conhecimento com foco na tecnologia está vinculada à análise dos instrumentos e ferramentas que podem mediar o processo de geração, retenção e reutilização da informação que sustenta o conhecimento pela organização e que por isso mesmo envolvem pessoas, metodologias e tecnologias da informação. A informação, conforme analisa CHOO (2003) é matéria-prima do cotidiano do ser humano e para isso ele tem os cinco sentidos que captam as mensagens do ambiente. Considera-se que a tecnologia da informação tem papel chave no processo, com utilização de estruturas *web* e portais corporativos, com tendência a forte crescimento, processos de educação à distância e estratégias de *data warehousing*, *data mining*, *web mining* e *text mining*. No contexto da presente tese, destacaremos os seguintes suportes à gestão do conhecimento que podem ser ofertados pela tecnologia da informação:

- Geração de informação estratégica através metodologias relativas ao tratamento da informação gerencial e suas tecnologias de implantação, como *data warehousing*;
- As aplicações de *software* oriundas de aplicação de propostas das áreas de descoberta de conhecimento (ou sistemas inteligentes) e recuperação da informação, tais como a mineração de dados e a mineração de textos;
- As aplicações de software oriundas do campo da computação colaborativa, tais como as ferramentas de correio eletrônico e workflow;
- As aplicações de gestão eletrônica de documentos, uma vez que apóiam a geração e administração das bases de dados textuais nas organizações e o conhecimento humano é mais fácil de ser expresso (codificado) na linguagem natural presente nos documentos;
- As aplicações de portais corporativos, uma vez que, conforme já citado, são software que permite configurar uma rede de computadores de uma organização (*intranet*) de forma dinâmica no sentido de integrar distintos repositórios de conteúdos de informação sob um mesmo ambiente de acesso a eles. Através de um portal, desde que esteja devidamente credenciado, o usuário tem acesso a quantos repositórios de dados e documentos que a sua organização possui.

7.2 Informação Estratégica no Suporte à Gestão do Conhecimento

Na medida em que evoluíam os sistemas de informação, a partir dos anos 80, no seio das organizações, as empresas foram percebendo o potencial da tecnologia para gerar uma visão delas mesmas e do ambiente em que se encontravam, buscando melhorar seu posicionamento estratégico. Entretanto, as informações que lastreavam essa visão estratégica não eram as informações transacionais, mas aquelas que foram gradativamente ganhando o nome de estratégicas ou decisórias, em um primeiro momento. Por informações estratégicas ou decisórias entendemos aquelas informações que não estão diretamente vinculadas ao controle rotineiro da atividade operacional da organização, mas que tem por finalidade subsidiar o posicionamento ou a escolha organizacional em decisões importantes[63]. São, portanto, informações das quais sobremaneira emergem o conhecimento necessário para definições estratégicas, geração

[63] Para uma análise minuciosa dessa questão, consultar TURBAN, E. et allli. *Decision Support Systems and Intelligent Systems* Upper Saddle River: Prentice Hall, 1998

de inovações e mudanças em rumos, processos e produtos da organização. Tais informações podem ser, por exemplo, indicadores sobre nível de satisfação dos clientes, históricos sobre comportamento de vendas de grupos de produtos por região e segmentadas por cliente ou tipo de cliente, informações sobre o comportamento de variáveis financeiras como custos diretos e indiretos por processo ou produto, comportamento de mercados dentre outros. E com o tempo, novas expectativas surgiram. Tais informações se distinguem do que podemos chamar de informações operacionais, ou do "dia-a-dia", como por exemplo o registro contábil de uma transação de venda a cliente ou compra de fornecedor, muito embora essa informação operacional possa ser "embutida" na informação estratégica ou decisória, ou seja: para se chegar a uma informação estratégica sobre comportamento de vendas numa região, por exemplo, foi preciso considerar uma base de dados que registrou as transações de vendas em diversos períodos e diversas regiões. Por isso as organizações passaram a considerar importante a adoção de metodologias e estratégias tecnológicas que possam viabilizar a identificação, desenho e extração do que chamamos de informação estratégica ou decisória das bases de dados das organizações, informações essas que, do nosso ponto de vista, se constituem num dos elementos motores da geração de conhecimentos que interessam a uma organização. É por essa razão que voltamos nossa atenção para Fatores Chave de Sucesso (FCS's) e *Balanced Scorecard* (*BSC)*, duas abordagens que selecionamos e consideramos relevantes nessa perspectiva no ambiente contemporâneo.

Conforme descreve FURLAN (1997), os FCS's descrevem aquelas poucas coisas que devem ocorrer de modo correto, mesmo em detrimento de outras que não estejam indo a contento. Esse autor também descreve a origem dessa abordagem:

> *" Fator Chave de Sucesso é uma teoria lançada como resultado da pesquisa pioneira de John F. Rockart do MIT - Massachusetts Institute of Technology em 1979. Desde então seu uso tem sido verificado em uma infinidade de projetos em todo o mundo como parte da metodologia de planejamento estratégico. A base dos estudos desenvolvidos pelos pesquisadores do MIT foi o principal executivo da organização; no entanto, estudos subseqüentes demonstraram que os fatores chaves de sucesso podem ser úteis para o nível intermediário de decisão, responsável por uma ou mais áreas funcionais. Os FCS's, se satisfatoriamente atendidos, garantem o desempenho competitivo da organização, mesmo se outros fatores menos relevantes forem negligenciados. Conseqüentemente, tais fatores deveriam receber cuidadosa atenção por parte da gerência executiva. A situação atual pode ser continuamente monitorada através da análise dos indicadores de desempenho. Poderíamos dizer que os fatores chaves de sucesso apresentam uma visão estratégica que considera não*

> *apenas o momento atual mas, principalmente, aspectos mais permanentes e abrangentes do negócio.* (FURLAN, 1997: 12-15).

A aplicação do método implica na identificação dos fatores chave, a partir dos objetivos da organização, e de indicadores de performance que permitem monitorar esses fatores chave. Esses indicadores são sustentados por sistemas que geram a informação que lhes corresponde.

Em termos de metodologias de planejamento de sistemas de tratamento da informação estratégica, surge na década de 90 o *Balanced Scorecard*[64], cuja a idéia central e viabilizar mais concretamente a ação estratégica. Pode-se tentar buscar visões que procuram conciliar aspectos das "duas escolas" de pensamento supra citadas. KAPLAN e NORTON (1997) irão apresentar a metodologia *BSC - Balanced Scorecard.*- ou Marcador Balanceado e que é definido para cada empresa de acordo com sua estratégia, constituindo um instrumento para gerenciar sua implantação. Na visão destes autores:

> "*O BSC traduz a missão e a estratégia das empresas num conjunto abrangente de medidas de desempenho que serve de base para um sistema de medição e gestão estratégica. O BSC continua enfatizando a busca de objetivos financeiros, mas também inclui os vetores de desempenho destes objetivos. O scorecard mede o desempenho organizacional sob quatro perspectivas equilibradas: financeira, do cliente, dos processos internos da empresa e do aprendizado e crescimento. O BSC permite que as empresas acompanhem o desempenho financeiro, monitorando ao mesmo tempo o progresso na construção de capacidades e na aquisição dos ativos intangíveis necessários para o crescimento futuro* (KAPLAN e NORTON, 1997: 2)".

Realmente, o desempenho de um negócio não pode ser medido em uma única dimensão. Por exemplo, o lucro de uma empresa é um aspecto importante do seu desempenho, mas sozinho não indica que a empresa está bem ou mal. KAPLAN E

[64] O *Balanced Scorecard* (marcador balanceado) pode ser considerado como uma metodologia de planejamento e definição de sistema de informação estratégica na medida em que procura apresentar uma série de conceitos e uma estrutura de organização dessa informação para medir o desempenho organizacional e subsidiar a tomada de posição da organização. Essa visão deve ser "balanceada", isto é: incluir não só a visão do resultado financeiro, mas também outras três dimensões importantes, conforme exposto nesse trabalho. É claro que, visto numa perspectiva mais ampla, não se reduz a metodologia, mas sua aplicação em conexão com sistemas de informação orienta o uso da tecnologia da informação e faz com que seu uso seja direcionado por sua perspectiva no desenho de sistemas.

NORTON (1997) desenvolveram o sistema de medição supracitado, conforme apresentado na figura abaixo.

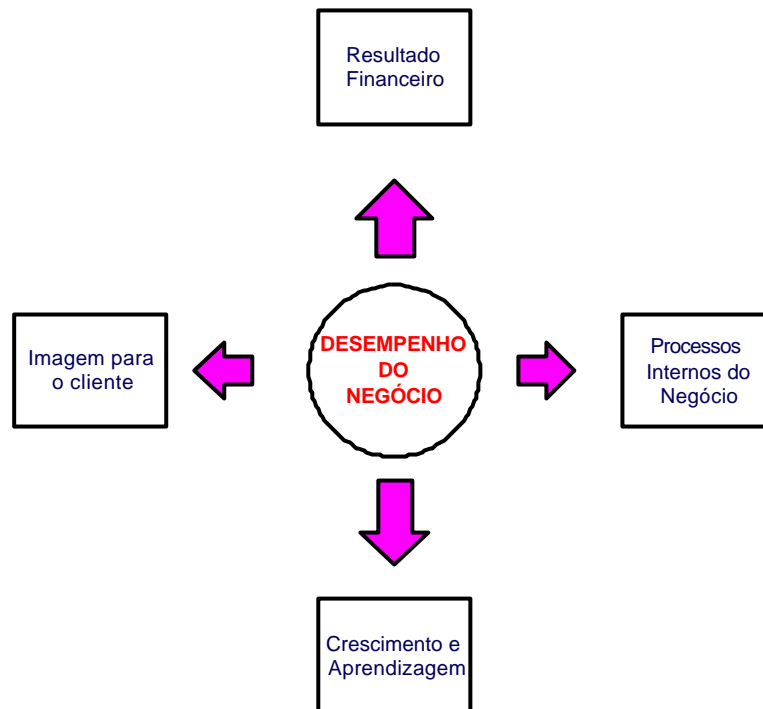



Figura 27: O *Balanced Scorecard* fornece a estrutura necessária para a tradução da estratégia em termos operacionais. Fonte: Adaptado de KAPLAN E NORTON (1997: 10)

Em cada uma dessas dimensões , são usados vários indicadores, relacionados às estratégias da organização. As quatro dimensões do desempenho compreendem:

- Resultados Financeiros: lucro, crescimento e composição da receita, redução de custos, melhoria da produtividade, utilização dos ativos e estratégia de investimentos;
- Posição em relação aos clientes: participação no mercado, retenção de clientes, aquisição de novos clientes, satisfação dos clientes, rentabilidade dos clientes, qualidade do produto, relacionamento com clientes, imagem e reputação;
- Processos internos do negócio: inovação (desenvolvimento de novos produtos e processos), operação (produção, distribuição vendas) serviços pós vendas (assistência técnica , atendimento a solicitações do cliente);
- Aprendizagem e crescimento: desenvolvimento de competências da equipe, infraestrutura tecnológica, cultura organizacional e clima para a ação.

Tanto os Fatores Chave de Sucesso quanto o *Balanced Scorecard* passaram a ser utilizados como referências conceituais para definição de sistemas de informação que visam atender necessidades estratégicas de informação, como demonstram a incorporação desses conceitos em produtos das empresas Microsoft, no produto SQL Server 2000 (Banco de Dados) e SAP, no produto My SAP mesmo sendo um ERP.

No campo do tratamento tecnológico da informação decisória ou estratégica, especialmente nos níveis gerenciais e técnicos mais altos na estrutura das organizações, as "tecnologias gerenciais" evoluem nos anos 90 para o conceito de *data warehousing*, forçada por suas crescentes necessidades de conhecimento, Por *data warehousing* entendemos um conjunto de novas tecnologias de armazenagem e geração de informação agrupando desde novos conceitos e abordagens em bancos de dados até novas visões de sistemas "inteligentes" que incorporam alguns recursos da chamada inteligência artificial. O uso da inteligência artificial (ou sistemas "inteligentes") no ambiente corporativo para tratar as bases de dados já existentes nas organizações fez surgir o conceito de mineração de dados ou *data mining*. Dedicaremos o próximo tópico ao ambiente *data warehousing* e ao *data mining*, tendo em vista sua relevância para a gestão do conhecimento e o crescente impacto que esse conjunto de tecnologia vem tendo no processo decisório das organizações conforme analisaremos a seguir.

7.2.1 *Data Warehousing* , Mineração de Dados e Mineração de Textos

Dentre as tecnologias que sustentam na organização o processo de gestão do conhecimento, podemos salientar os processos de *data warehousing*, *data mining e text mining*. LANCASTER (2004) faz uma critica severa ao uso das expressões *data mining*, *text mining* e *data warehouse*, afirmando que as mesmas merecem ser desprezadas, alegando que tratam-se de estratégias tecnológicas já definidas no passado como "descoberta de conhecimento". E tem razão, pois já há mesmo linhas de pesquisa e estudo que usam a expressão "descoberta de conhecimento". O autor da presente tese concorda com as afirmações de LANCASTER (2004) ponderando, entretanto, que tais expressões já se difundiram enormemente, não só no campo comercial, mas também acadêmico e por isso serão utilizadas, porém com tais ressalvas[65]. Muitas vezes a

[65] Ver: Uma Nota Sobre Terminologia (e a redescoberta da roda) *in*: LANCASTER. (2004): p. x-xiv. Esse autor afirma que a atividade de mineração, em seu sentido literal, jamais poderia ter sido usado

própria pesquisa sobre tais assuntos é impossível caso não se use a terminologia utilizada comercialmente e que foi absorvida também em grande extensão pela pesquisa acadêmica. Além disso, é importante salientar que, no caso da implementação de um "*data warehouse*" não se usam, em geral, modelos e técnicas oriundas da área de inteligência artificial, como é o caso dos processos tecnológicos denominados originalmente "descoberta de conhecimento". Também é importante salientar que, na literatura da área de Banco de Dados, *Data warehousing* e *data warehouse* são termos que representam realidades diferentes: o primeiro termo representa um conjunto de tecnologias que inclui bases de dados e softwares de tratamento e extração de dados. O segundo termo representa apenas um dos elementos desse conjunto que é a base ou armazém de dados, mas que isoladamente, acaba não tendo muito significado e sentido prático. Abordaremos nas seções 7.2.1.1 e 7.2.1.2 as idéias de *data warehousing* e mineração de dados, respectivamente. No que tange à tecnologia de mineração de textos, ou *text mining* (que pode também ser vista como um tipo específico de *data mining* para textos), realizaremos sua abordagem no capitulo 8. Essa visão panorâmica é importante uma vez que essas tecnologias são, em geral, utilizadas em torno de um mesmo eixo: a inteligência e o conhecimento embutido em dados e documentos. Por isso, no final dos anos 90, um novo termo surgiu para caracterizar tal eixo nas organizações: o de *business intelligence*.

7.2.1.1 Data Warehousing – Processo de Geração de Informações Estratégicas

Os sistemas de informações gerenciais e/ou *softwares* de aplicações estratégicas passaram a utilizar recursos de *modelagem multidimensional* embutidos na visão de *data warehousing* para atender adequadamente às proposições dos modelos dos fatores chaves de sucesso e do balanced scorecard, acima descritos. *Data warehousing* (literalmente: armazenagem de dados, mas com um sentido de guarda seletiva de informações previamente definidas como relevantes para processos decisórios) é uma estratégia tecnológica que significa organizar de maneira inteligente bases de dados organizacionais de forma a permitir que o processo decisório seja "*otimizado*" por uma

---

como analogia aos processos tecnológicos aqui descritos. O autor critica também o uso das expressões "taxonomia" e "ontologia" as quais substituíram a palavra "classificação" ou "esquema de classificação" dependendo do trabalho em que aparecem. Lembra que alguns autores execraram a substituição de classificação por ontologia e também critica o fato da palavra ter sido alvo de substituição por "categorização".

adequada classificação das informações a partir da separação entre dados estratégicos e operacionais. Um *data warehouse* é um banco de dados organizado de forma específica para atender às necessidades de informação estratégica e gerencial. *Data warehousing* é um "ambiente de transformação dos dados em informação" [66], através de metodologias específicas e utilização de ferramentas de extração, manipulação e exploração. É resultado de uma convergência de diferentes tecnologias, tais como bancos de dados, interfaces gráficas, computadores com alto poder de processamento e/ou processamento paralelo, sistemas operacionais de rede, discos rígidos com grande capacidade de armazenamento e velocidade acesso e, especialmente, o modelo *OLAP* (*Um Lenhe Analytical Processing*), etc. tornando viável o melhor uso de informações geradas nos diversos sistemas transacionais, no *ERP*, ou obtidas externamente. O ponto central é a idéia de "armazém de informações", ou *data warehouse*, uma base específica para informações relevantes ou estratégicas para a organização. Vários termos técnicos surgiram associados a esse, ganhando espaço e interesse nas organizações, tais como *OLAP*, d*ata mining*, modelagem dimensional, dentre outros. Portanto, *data warehousing* é um processo que é suportado por um número de diferentes produtos de tecnologia e serviços profissionais de tecnologia de informação

Conforme INMON (1996) [67] um *data warehousing* é uma área de armazenamento de informação para suporte a decisão, ou, num sentido mais amplo, de informação geradora de conhecimento sobre operação, produção, resultados, produtos, mercados, ou quaisquer variáveis consideradas estratégicas para a organização. Ele coleta dados a partir de diversas aplicações em um sistema operacional de uma organização, integra os dados em áreas lógicas de assuntos dos negócios, armazena a informação de forma que ela fique acessível e compreensível a pessoas não técnicas responsáveis por tomadas de decisões e entrega informação aos tomadores de decisão através da organização e definição de várias ferramentas de relatório e busca. Na perspectiva desse autor:

---

[66] No sentido que coleta, articula, armazena e disponibiliza dados já existentes em outros sistemas da organização e conferindo a eles um novo desenho, adequado ao processo de tomada de decisões, processo esse fortemente estudado por CHOO (2003).

[67] Esse autor é o pioneiro na definição de conceitos e metodologias nessa área, no início dos anos 90. Ver INMON, William H. Buiding the data warehouse. John Wiley & Sons, Inc. USA, 1996

- Os dados são organizados por área de assunto: Áreas de assunto representam uma coleção de todos os dados na organização que pertençam a um tópico que é considerado importante para um tomador de decisão. Por exemplo, o assunto "trens" pode conter toda informação disponível a partir de diferentes sistemas operacionais sobre a história do movimento das máquinas locomotivas e carros ao longo de um sistema ferroviário.

- Os dados são integrados: Os dados devem ser transformados em formatos comuns de medida, referência e armazenamento para que se tornem úteis. Por exemplo, uma companhia de seguros pode ter informação sobre diferentes apólices de um mesmo agente, armazenada em diversos bancos de dados, utilizando tecnologias radicalmente diferentes. Para tomar decisões efetivas sobre o relacionamento total com os clientes, os dados precisam ser apresentados em um formato comum. Além disso, se a companhia está para tomar decisões referentes a lucros, ela deve concordar em regras comuns de negócios tais como a mensuração da lucratividade.

- O "*warehouse*" não é volátil: A informação é carregada no *warehouse* e então acessada para a tomada de decisão. Isto contrasta com um sistema operacional que é atualizado tão logo um novo evento ocorra. Então, por exemplo, o *warehouse* não será modificado tão logo um cheque for compensado no banco. Essa mudança será refletida no carregamento subseqüente da informação no *data warehouse*, quando será então detectada por um analista que estiver procurando por todos os relatos que indiquem uma tendência particular.

- A informação é orientada pelo tempo (dimensão básica e comum em toda modelagem): O *data warehouse* é uma seqüência de "*snapshots*" (instantâneos – como se fossem "fotos" – de um determinado momento específico) de informação organizacional tiradas em intervalos de tempo pré-definidos, tais como todos os dias ou uma vez por semana. Isso significa que, por exemplo, um analista de vendas em uma fábrica de chocolate pode analisar três anos de histórico de vendas com "*snapshots*" de informação que foram coletadas no final de cada mês ou semana.

Então, por *data warehouse* podemos entender uma modalidade de implementação de base de dados informacional computadorizada voltada ao armazenamento de dados compartilháveis, obtidos a partir dos ambientes de base de dados operacionais. Trata-se de uma base de dados temática, ou por assuntos, que permite aos usuários penetrar o vasto repositório de dados operacionais de uma corporação para subsidiar mais eficazmente um processo decisório e/ou fazer frente às tendências de negócios, facilitando esforços de planejamento. Seu princípio fundamental afirma que as bases de dados nos sistemas de suporte à decisão ou qualquer sistema que trate informações estratégicas e mais refinadas, devem ser isolados dos sistemas transacionais. Tais bases devem desenhadas especificamente para finalidades estratégicas e de tomada de decisão. Assim, podem ser alimentadas pelos sistemas transacionais - mas não utilizadas por eles - e seu desenho deve privilegiar o desempenho das consultas. É uma tecnologia de armazenamento estratégico de informação e que vem ganhando cada vez mais espaço dentro das organizações e que pode ser utilizada na construção de bases de dados de conhecimento coletivo. Essa base de dados é constituída com uma abordagem diferente dos bancos de dados relacionais, cujo princípio fundamental é o relacionamento de múltiplas tabelas entre si[68]. Diferentemente, no *data warehouse* a modelagem de dados é chamada de multidimensional, ou seja: ao invés de considerar tabelas, considera dimensões de informação. Em sua obra clássica sobre o assunto, KIMBALL (1996) afirmará:

> *"Modelagem dimensional é um nome novo para uma técnica antiga usada para criar bancos de dados simples e compreensíveis. Quando um banco de dados pode ser visualizado como um "cubo" contendo três, quatro, ou até cinco ou mais dimensões, as pessoas conseguem fatiar esse cubo em qualquer de suas dimensões. (...) Experimentemos um exemplo simples. Imagine um negócio em que o CEO descreva as atividades da empresa da seguinte forma: "vendemos Produtos em vários Mercados e avaliamos nosso desempenho ao longo do Tempo". Muitas pessoas consideram fácil imaginar esse negócio como um cubo de dados, com legendas em cada uma das arestas do cubo, como mostra a figura abaixo. Qualquer ponto no interior do cubo está na interseção das coordenadas definidas pelas arestas do cubo. Para o exemplo citado, legenda mos as margens do cubo como Produto, Mercado e Tempo (KIMBALL, 1996:Introdução p. xviii)".*

[68] Assunto já abordado no Capitulo 5. Na grande maioria das organizações contemporâneas, o uso de bancos de dados vem seguindo os princípios do modelo relacional proposto por E.F CODD (1969) e que consiste de múltiplos relacionamentos entre diversas tabelas, criadas através de sistemas gerenciadores de bancos de dados e que se sevem de implementações a partir da técnica da análise de entidades e relacionamentos, a qual não é objeto de estudo direto dessa dissertação. A abordagem multidimensional, aqui especificada, é uma alternativa proposta a essa técnica, para tratar informações gerenciais e estratégicas, mas sem excluí-la completamente, tendo em vista que a maioria das organizações atualmente continua a operar com bancos de dados relacionais.

A figura 28 abaixo, adaptada da obra de THOMSEN (1997) e de KIMBALL (1996) apresenta a idéia desses autores sobre modelo multidimensional e o processamento analítico em tempo real (*On Line Analytical Processing* - OLAP). Nessa figura, por mercado entenda-se a região na qual o mesmo se encontra subdividido. Conforme lembra esse autor, Isso permite que os dados sejam separados e combinados usando-se qualquer medição possível do negócio.

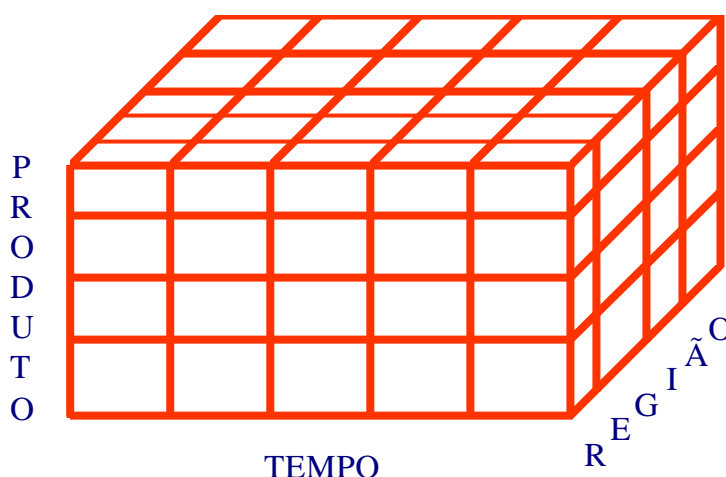



Figura 28: O modelo multidimensional de um negócio: cada ponto do cubo contém as medições para a combinação específica de produto, mercado e região.
Fonte: Adaptado de KIMBALL (1996: Introdução) e THOMSEN (1997: 83)

Entretanto, como a maioria dos bancos de dados nas organizações é relacional, a implementação de sistemas de informação, voltados para descobrir conhecimento com valor estratégico (para apoiar decisões), com base nessa abordagem tornou-se bastante complexa. Realmente, a maioria dos sistemas das organizações continua no ambiente transacional, chamado de OLTP - *On-line Transaction Processing* (processamento de transações on-line) responsável por registrar e manipular dados de transações cotidianas de uma organização. Uma transação é um evento completo com uma ou múltiplas seqüências relacionadas. Por exemplo, uma compra em um supermercado terá vários impactos nos sistemas e bancos de dados: registro no sistema de faturamento, registro de retirada em estoque e contabilização. É nesse ambiente que se encontram todos os sistemas e aplicações direcionadas ao dia-a-dia operacional da empresa. Esse ambiente opera com dados excessivamente detalhados, isto é, organizados por transação e por

isso é inadequado aos sistemas de apoio a decisão. Novamente KIMBALL (1996) aponta:

> *"a despeito de todas essas diferenças, continuamos a usar os princípios e ferramentas OLTP para projetar nossos bancos de dados de data warehouse. (...) As técnicas e padrões de projeto apropriados para o processamento de transações são inadequadas e até mesmo destrutivo para projetos de data warehouse*" (KIMBALL, 1996: Cap1)".

A questão se apóia basicamente na modelagem de bancos de dados relacionais, até então realizada no chamado modelo de entidade-relacionamento e voltada para o ambiente OLTP. Mas o problema vem sendo solucionado não tanto com novos softwares de gerenciamento de bancos de dados, mas com artifícios metodológicos para modelagem de dados[69]. Como a grande maioria das organizações possui grandes bases de dados operacionais modeladas segundo no ambiente relacional, urgia criar-se uma abordagem metodológica que viabilizasse a idéia da modelagem por dimensões usando, porém o próprio SGBD relacional existente numa determinada organização, o que se consolidou com o modelos como o estrela – *star model*.

O modelo estrela assume que existem apenas dois tipos de tabelas: as de dimensão e de fatos. De acordo com NETO (1998), o diagrama estrela é a principal tipo de modelo multidimensional. Nele, há uma tabela central, dominante, com muitas junções, conectando-se a outras tabelas, as tabelas de dimensão, as quais possuem apenas uma junção com a tabela central. As tabelas de "fatos" tem característica agregadora, promovendo a articulação das dimensões. As tabelas de dimensão não podem relacionar-se entre si, mas apenas com a tabela de fatos. Uma tabela de fatos armazena medições numéricas do negócio. Cada uma das medições é obtida com a

[69] Para viabilizar tecnicamente a implementação de *data warehouses,* numa perspectiva muldimensional, mas considerando que a maioria das bases de dados das organizações estão implementadas em sistemas gerenciadores de bancos de dados – SGBD's de natureza relacional, foram desenvolvidos processos metodológicos para evitar que as organizações não precisassem abandonar seus SGBD's relacionais, como por exemplo os definidos pelos uso de softwares *ORACLE®, SYBASE®, IBM DB2®* ou *Microsoft SQL SERVER®*. Surgiram então, dentro da perspectiva de modelagem multidmensional, as técnicas "modelo estrela" – *star model* (com suas variações *galaxies* e *constellations schemas*) e modelo "floco de neve" – *snowflake model*. Para uma visão mais detalhada desse assunto, que abordaremos aqui superficialmente, ver o artigo Mapeamento entre os modelos E/R e Star. NETO, Roberto DCC/NCE/UFRJ. Rio de Janeiro, 1998 e KIMBALL (1998). É bom também pontuar que surgiram, ao longo da década de 90, softwares gerenciadores de bancos de dados multidimensionais que não se consolidaram no ambiente corporativo, justamente pelo grau intensivo de uso dos SGBD's relacionais que deviam, então ser ajustados para um novo fim: gerar informação gerencial através da implementação de modelos como estrela e floco de neve.

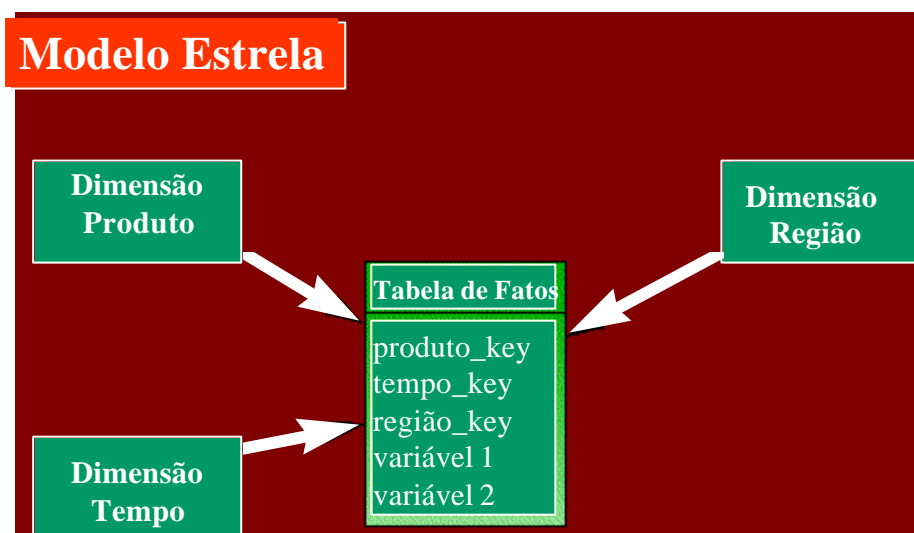
interseção de todas as dimensões. A figura 29, abaixo, detalha um exemplo que começou a ser discutido na Figura 29 na apresentação da modelagem multidimensional. O modelo estrela tem grande poder de visualização de informações quando comparado com o modelo relacional. Evita redundância de dados e campos de tabelas e promove uma melhor capacidade de percepção e agregação de informações. É, portanto, o grande recurso de ordem metodológica que permitiu a expansão da área de *data warehousing* nas organizações.

- **Dimensões: possíveis formas de visualizar os dados. Por exemplo, tempo, região, produto**
- **Variáveis: medidas numéricas. Por exemplo, quantidade estocada, vendas, lucro**
- **Tabela de Fatos: Tabela central do modelo estrela que permite a associação entre dimensões**

Figura 29: Modelagem Multidimensional de Dados Para o Processo Decisório no Modelo Estrela
Fonte: Adaptado de KIMBALL (1996: 27-29)

7.2.1.2 Mineração de Dados (Data Mining)

Além dos sistemas de informações gerenciais que utilizam estratégias de modelagem multidimensional, outro recurso de prospecção em bases de dados é a descoberta de conhecimento em bases de dados ou mineração de dados, já analisada no tópico 6.4 da presente tese. A abordagem de d*ata mining* dispõe de várias técnicas para a execução dessas tarefas, tais como tratamento estatístico, algoritmos genéticos, lógica *fuzzy* e redes neurais (apresentadas no tópico 6.3 desse trabalho).

Quando determinados padrões de comportamento, como associação de produtos durante um processo de compras, por exemplo, começam a se repetir com freqüência, as ferramentas *data mining* indicam a presença de oportunidades e "*Insights*" em relação àquele público consumidor. O diferencial de processos de *data mining* está no fato de que as descobertas de padrões se dão por uma lógica de algoritmos com base, por exemplo, em uma rede neural. São ferramentas de descobertas matemáticas feitas sobre os registros corporativos já processados, muitas vezes em cotejo com descobertas empíricas.

Por exemplo, o estudo do perfil dos clientes de uma empresa é de extrema importância para que os seus analistas de negócio possam desenvolver *mix* de novos produtos e novas estratégias de relacionamento mais eficazes e de melhor retorno. Tal estudo permite a identificação de padrões no comportamento dos clientes, determinando, por exemplo, sobre quais clientes determinado tipo de estratégia de *marketing* possui melhores resultados, ou então quais são os prováveis compradores de um novo produto, em que condições e em que perspectivas de fidelização.

As técnicas de *data mining* permitem que esta análise seja realizada usando os bancos de dados da empresa, extraindo valiosas informações que estão presentes nos dados mas que ficam ocultas devido ao grande volume de informação. É importante ressaltar que nenhuma alteração na plataforma de *software/hardware* da instituição precisa ser feita para que se possa construir um sistema de *data mining*. Independentemente da forma como estes dados estão armazenados eles podem ser utilizados para alimentar o sistema que pode ser facilmente acoplado aos demais sistemas da empresa.

Em recente trabalho, WEIXUAN, ZHENGXIN & YONG (2004)[70] *et alli* apresentaram estudos associando o uso da mineração de dados com a gestão do conhecimento, analisando suas metodologias, casos práticos de integração dessa tecnologia com a gestão do conhecimento organizacional. Tais estudos constituem uma visão avançada desse vínculo e do processo de descoberta de conhecimento em bases de dados. Discutem a relevância e aplicabilidade na análise e classificação de dados

---

[70] Anais do Chinese Academy of Sciences Symposium on Data Mining and Knowledge Discovery – CASDMKD realizado em Pequim, na China, em 2004. Ver especificação completa nas referências.

biomédicos, técnicas de agrupamento (*clustering*), análise, prospecção e visualização dos dados, sistemas de conhecimento empresariais e o processo de descoberta de conhecimento como um todo a partir de bases de dados.

7.3 Computação Colaborativa em Gestão do Conhecimento: Trabalho Cooperativo Suportado por Computador (CSCW – *Computer Supported Cooperative Work*)

A integração de pessoas e organizações através de computadores é um processo que vem se aprofundando nos últimos anos e ganhou novo impulso pelo crescimento do uso das redes de computadores, mormente da Internet[71]. Nas empresas, além da *Internet*, que tem uso crescente, as redes internas, ou *intranets*, também fornece sua parcela de contribuição para ampliação do uso de computadores pelas pessoas. Os sistemas computadorizados podem também viabilizar o trabalho colaborativo em grupos ou comunidades virtuais – *groupware*. A idéia de *groupware* ganha impulso na medida em que cria novas perspectivas de relacionamento e trabalho dentro das empresas.

De acordo com BARTHELMESS (1996):

> *"a área de CSCW compreende todo o software que tem por objetivo prestar auxílio ao trabalho cooperativo. É justamente este o fator que torna este tipo de software diferente dos demais.*
>
> *A noção de que este tipo de software deve mediar a interação de diversas pessoas que buscam obter um objetivo comum introduz novos requisitos normalmente não encontrados em outros sistemas. Se enfatiza a interação entre usuários e não mais a interação sistema/usuário, como acontece na maioria dos sistemas (que passaremos a chamar de convencionais, em contraposição a esta nova classe de software cooperativo ou groupware).*
>
> *A ubiqüidade de estações individuais de trabalho, ligadas por redes, cria uma oportunidade tecnológica de se prover suporte às atividades de grupos de trabalho nas organizações . O padrão de utilização dos computadores migrou de uma centralização representada pelas máquinas de grande porte nos CPDs para um de utilização individual, em que cada usuário ou grupo de usuários possui suas ferramentas, como planilhas e editores, trabalhando normalmente em isolamento. Estas atividades isoladas não correspondem, porém, à real necessidade das organizações, nas quais o trabalho não é realizado costumeiramente por apenas um indivíduo, mas é fruto de um esforço coletivo.*

---

[71] Para um estudo aprofundado sobre redes de computadores ver: SOARES, Luiz F. G. et. al . Redes de Computadores: Das LANs, MANs e WANs às Redes ATM. Rio de Janeiro. Campus, 2005

Diversas categorias de produtos procuram explorar esta nova oportunidade, geralmente em áreas específicas, como por exemplo os da lista abaixo, que está longe de ser exaustiva:

- *Editores e planilhas para uso em grupo;*
- *Vídeo-conferência;*
- *Bulletin-Board Systems;*
- *Correio eletrônico;*
- *Sistemas de Workflow* (BARTHELMESS,1996: Cap 1)*".*

Merecem destaque no contexto do *groupware* o correio eletrônico e o *workflow automatizado, ou simplesmente workflow* (fluxo de trabalho desenhado para ser realizado de modo automatizado). O correio eletrônico é um recurso mais antigo que já se encontrava presente em ambientes de computação centralizada com uso de *mainframes,* mas é com a Internet e uso de ambiente cliente-servidor que ganhará uma maior e mais profunda amplitude de uso nas organizações. A troca de mensagens eletrônicas evolui gradativamente para o conceito de *workflow*, ou fluxo de trabalho computadorizado, isto é, a incorporação de rotinas e processos encadeados de trabalho por *softwares* em computadores ligados em rede**.** Podemos ter a utilização de *workflow automatizado* para geração e revisão de documentos técnicos

Ainda de acordo com BARTHELMESS (1996):

Sistemas de workflow podem ser definidos como sistemas cujo objetivo é "auxiliar as organizações na especificação, execução, monitoramento e coordenação do fluxo de trabalho em um ambiente de escritório distribuído"

Software especificamente desenvolvido para possibilitar a criação de sistemas de gerenciamentos e de compartilhamento de informações por multi-usuários, possibilitando a consulta em rede em tempo real. Este sistema pode ser moldado de acordo com as necessidades de cada empresa

O fluxo de trabalho corresponde à tramitação de documentos e informações entre diversos agentes, pertencentes a unidades organizacionais potencialmente distintas, cada qual agregando uma parcela do trabalho necessário para que se alcance determinado objetivo de negócio.

A observação de que por volta de 90% desta tramitação pode ser considerada rotineira e repetitiva torna natural que se deseje suprir suporte automatizado para estas tarefas. Após um surto de interesse nos anos 70, associado à área de automação de escritórios, a pesquisa relativa a sistemas de workflow ficou estagnada durante a maior parte dos anos 80, devido ao fracasso dos sistemas pioneiros. Verificou-se que estes sistemas apresentavam uma inflexibilidade que inviabilizava seu uso prático, mesmo em situações simples e controladas (BARTHELMESS, 1996: Cap 2).

Na perspectiva de aumento da produtividade e da flexibilidade nas empresas, a idéia do "teletrabalho" se afirma mais em meados da década de 90. Ressurge, portanto, a pesquisa na área, com o aparecimento de sistemas mais flexíveis, baseados na experiência de seus predecessores de primeira geração. Surgem *softwares* que integram as funções de fluxo eletrônico de documentos em processos empresariais e correio eletrônico, como o *Exchange*®, da Microsoft® e o *Lotus Notes*®, da IBM-Lotus®. Na visão dos fabricantes, o principal objetivo de softwares de *workflow* é aumentar a eficiência de processos de negócio e a eficiência das pessoas que trabalham em conjunto para executá-los. As tecnologias para *groupware* oferecem ferramentas cooperativas como fóruns de discussão, calendários, agendas e um sistema de distribuição de arquivos. Os benefícios em geral ressaltados quando do uso desses recursos são:

- Correio eletrônico
- Gerenciamento de agenda
- Possibilidade de acesso a documentos por vários usuários simultaneamente
- Velocidade de localização de documentos
- Automação relativamente fácil de rotinas administrativas e processos de trabalho
- Redução de espaço físico devido a eliminação de papéis
- Análise de projetos por diversas pessoas sem necessidade de estar fisicamente presentes
- Uso de teleconferências.

Analisando um *software* voltado para o ambiente de *groupware,* o *Lotus Notes*®, percebemos que é um sistema de desenvolvimento de aplicações, personalizável e programável, que apresenta uma boa integração com a Internet. O *Lotus Notes* proporciona às organizações um interessante conjunto de serviços baseados em padrões mais atuais, especialmente para correio eletrônico, grupos de discussão e uso de calendário partilhado.

O *Notes* introduz o conceito de ambiente de *groupware* total, em oposição a um conjunto de funções rigidamente definidas e independentes. As aplicações de *groupware* do *Lótus Notes* - desde as aplicações de *workflow* até às bases de dados de

discussão - são construídas dentro desse ambiente utilizando blocos de construção comuns. A capacidade para especificar e desenvolver uma base de dados adequada a uma aplicação particular de uma empresa é essencial para muitos administradores de *groupware*. No entanto, para instalações mais pequenas, com menores recursos de desenvolvimento, o Notes também oferece um conjunto de modelos de aplicação, incluindo uma visão básica de *workflow* com aprovação por item, bibliotecas de documentos, um sistema de reserva de espaço e uma base de dados de discussão.

Reconhecendo a importância do conhecimento tácito, poucas posturas tecnológicas são tão eficazes em termos de apoiar sua geração como as estratégias de trabalho em grupo e trabalho colaborativo por computador, visando o estímulo da interação entre pessoas e com a realidade e faz liberar novos conhecimentos através da resolução de problemas e da relação com o trabalho compartilhado.

7.4 Gestão Eletrônica de Documentos – GED

Nas organizações em geral os documentos são gerados de forma intensa e seu controle em geral difícil, sobretudo no que tange à organização e recuperação, tanto os documentos físicos quanto os digitais. Diferentes e às vezes conflitantes processos de classificação, indexação, armazenagem e recuperação se verificam. Quando há dispersão geográfica de uma empresa a situação piora. Muitas vezes, há dificuldade em se recuperar um documento. A existência de um acervo, ainda que amplo, não garante a geração da informação como suporte à gestão do conhecimento. São problemas constantes das organizações:

- Sistemas de recuperação de informações ineficientes e às vezes diferentes.
- Vários ambientes
- Controles paralelos departamentais.
- Experiência do quadro de pessoal da equipe.
- Não padronização dos dados.
- Demora na pesquisa.
- Pouca visibilidade/transparência do acervo documental.
- Não controle do conteúdo.

De acordo com SILVA (1995), sistemas de gestão eletrônica de documentos, são

> *"sistemas desenvolvidos para o gerenciamento de todo o ciclo de vida de um documento, desde a sua geração, manutenção (alterações, inclusão de anotações, dentre outras), guarda, pesquisa e recuperação, até o seu descarte" (p.9).*

GED é um conjunto de tecnologias que permite o gerenciamento de documentos de forma digital. Tais documentos podem ser das mais variadas origens e mídias, como papel, microfilme, som, imagem e mesmo arquivos já criados na forma digital. Baseando-se nas informações do CENADEM (2006), podemos dizer que antigamente, a tecnologia de GED enfatizava basicamente a digitalização de um documento gerado em papel através de um *scaner*. Assim, ele poderia ser visualizado na tela do computador, inclusive em rede. Hoje, entretanto, a maior parte das organizações tem uma imensa quantidade de documentos que já gerados eletronicamente: são os documentos de editores de texto, planilhas eletrônicas, desenhos de engenharia, e-mails. Esse novo acervo de documentos digitais, sem gerenciamento, implica em muitas cópias de diversas versões, duplicação de arquivos, provocando a má utilização de espaço em discos, e custos desnecessários com armazenamento. Para administrar esse contexto digital, quer sejam ou não originalmente eletrônicas, o conceito sobre o que é GED ampliou-se como que automaticamente. Gerenciar documentos é hoje quase como que cuidar de toda a vida informacional da empresa.

Conforme já afirmado, as aplicações de gestão eletrônica de documentos apóiam a geração e administração das bases de dados textuais nas organizações uma vez que o conhecimento humano é mais fácil de ser expresso (codificado) na linguagem natural presente nos documentos textuais, sejam textos livres ou com algum tipo de estrutura. A GED pode e deve se revelar como um instrumento viável para dinamizar, racionalizar e disseminar o acesso à informação e a tomada de decisão no nível estratégico das organizações.

A GED tem o desafio de lidar com as seguintes tarefas:

- Catalogação/indexação, que implica na definição das características que identificam cada tipo de documento/formato, e estabelecimento da estrutura do banco de dados e criação de interfaces para entrada/atualização de dados.

- Pesquisa e recuperação – talvez a função mais relevante da GED e que deve viabilizar a rápida recuperação de um documento, ou conjunto de documentos armazenados nos bancos de dados, com o mínimo de informação específica (parte de nomes, descritores, datas aproximadas, dentre outras).

- Gerência do fluxo de documentos, que envolve a definição e operacionalização de seu fluxo ao longo das redes de computadores interna e externa da organização, levando-se em conta questões de sigilo, segurança, roteiros, autorizações níveis e autorizações tipos, anotações e associação de outros documentos.

- Gerência de versões dos documentos, estabelecendo controles sobre as distintas versões de um mesmo documento, evitando-se que documentos em versões mais antigas sejam tratados como atuais. É possível manter cópias completas de todas as versões dos documentos, ou então manter a versão original e o histórico das modificações introduzidas, dependendo do *software* de gerenciamento. Atualmente, muitos editores de textos possuem mecanismos que permitem o gerenciamento de versões.

- Gerência do uso corporativo e individual, com instalação de rotinas de *backu*p, mecanismos de segurança e controle de acesso (catalogação de usuários, grupos, níveis de acesso e utilização de recursos), manutenção de índices, definição dos procedimentos da rotina operacional do sistema e integração da GED com outros sistemas e controle estatístico de utilização e alocação de recursos.

O conceito de GED vem evoluindo ao longo dos anos. Um conceito oriundo dessa evolução é o de *Enterprise Content Management* é o conjunto de tecnologias utilizadas para captar , gerenciar , armazenar , preservar e disponibilizar conteúdo e documentos relativos aos processos organizacionais. As ferramentas e estratégias de ECM permitem o gerenciamento das informações não-estruturadas da empresa , onde quer que tais informações se encontrem. O CENADEM (2006) apresenta, na figura 30, o seguinte fluxo operacional para o processo de ECM/GED:

CAPTAÇÃO	ARMAZENAMENTO	GERENCIAMENTO	DISTRIBUIÇÃO	PRESERVAÇÃO
É o modo como o informações e documentos, eletrônicos ou em papel, passam para um repositório de conteúdo para reutilizar, distribuir e armazenar	É onde se coloca o conteúdo e como você o encontra novamente.	São as ferramentas e técnicas para mover conteúdo pela organização e monitorar o desempenho delas.	Como você consegue o conteúdo certo para o público certo no dispositivo certo. (Acesso)	As suas opções para armazenamento a longo-prazo do conteúdo essencial de sua empresa.

Figura 30: Fluxo operacional da Gestão Eletrônica de Documentos
Fonte: CENADEM (2006)

As tecnologias de GED são a importantes para o estabelecimento de programas de Gestão do Conhecimento, já que elas gerenciam o conhecimento existente na empresa que foi explicitado sob a forma de documentos. E é necessário que o documento e a informação sejam registrados e distribuídos de forma adequada na organização para o atingimento dos objetivos que ela possui. O documento sempre estará atrelado ao compartilhamento do conhecimento, à tomada de decisões, concretização de negócios, processos e métodos internos e outras importantes ações. Como as organizações ainda geram grande quantidade de documentos em papel, deverão também criar estratégias para gerenciá-los de forma eficiente e rápida. Para a adequada implantação de estratégias de gestão do conhecimento é necessário o registro de todos os documentos que entram e saem da empresa, com inclusões, alterações e conteúdos. E isso só é possível com o uso de tecnologias de GED.

7.5 Portais Corporativos

Com o passar do tempo, a maioria das organizações atravessou um caminho de crescente informatização, criando diversas aplicações que começaram gradativamente a clamar por integração recíproca, uma vez que a integração de processos e fluxos de informações conduz a um melhor desempenho. Esse processo de integração, que se amplia a cada dia na chamada economia digital é hoje suportado por três tipos de tecnologia, cada uma com suas características e funcionalidades:

- *ERP – Enterprise Resources Planning* (já definido no início desse capítulo)
- *EAI – Enterprise Application Integration* – Integradores de Aplicações Corporativas
- *EIP – Enterprise Information Portals* - Portais de Informações Empresariais

Definiremos a seguir os conceitos de *EAI* e *EIP*.

7.5.1 – Integradores de Aplicações Corporativas

A idéia de *EAI* pode ser compreendida como um conjunto de "pontes" para interligar os vários sistemas de informação, tanto os adquiridos de diferentes fornecedores quanto os desenvolvidos internamente. Diante dos muitos sistemas, inclusive o *ERP*, operando em uma organização, o ideal seria que se criasse um único sistema que integrasse todos os outros numa interface única que permitisse o trânsito de usuários autorizados por eles. Este novo sistema é classificado como *EAI.* A figura 31 a seguir ilustra a passagem de uma integração sem *EAI* e com *EAI.*

Conforme LINTHICUM (2000) as tecnologias de EAI são sustentadas por elementos que permitem tanto a um banco de dados como a um software aplicativo se comunicar com outros bancos de dados ou softwares aplicativos, eliminando os níveis de divergência entre o que precisa ser integrado. Esses elementos, que atuam tanto no nível lógico quanto físico, são definidos por *middleware*[72].

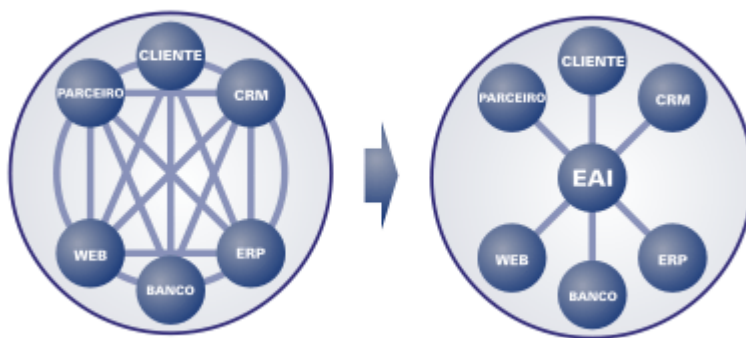



Figura 31: Processo de integração entre aplicações – migrando para um ambiente com EAI.
Fonte: IDEAIS Consultoria – www.ideais.com.br. Acessado em 22 nov. 2006

Por outro lado, vale lembrar que nos últimos dez anos o ambiente *web* teve um crescimento significativo na sociedade e percebemos uma explosão de redes internas

[72] Eles incluem, por exemplo, os objetos distribuídos.

baseadas em sua tecnologia na empresas. Constatamos também a ampliação de estruturas externas de relacionamento com clientes e fornecedores, baseadas nesse ambiente. Outros fatores, como educação à distância e entretenimento pressionam o crescimento da Internet no mundo contemporâneo e seu número de usuários continua crescendo. Há cada vez mais aplicações *web* e o número de documentos nesse ambiente é difícil de calcular. Muitas vezes um potencial usuário fica perdido diante de tantas possibilidades.

Considerando, tanto a proliferação de muitas aplicações de software no ambiente empresarial, a crescente disponibilidade de conteúdos em *intranets* e na *Internet*, tornou-se necessário o advento de algum tipo de recurso tecnológico que pudesse facilitar a vida do usuário em sua relação com a tecnologia da informação, usando a interface do ambiente WEB e esse ambiente é o portal. Em uma heterogeneidade de plataformas de hardware e software e complexidade crescente dos ambientes tecnológicos das empresas, *ERP* e *EAI* permitem em geral que as aplicações falem entre si, mas essas aplicações precisam falar também com o usuário, o que é papel do *EIP*.

7.5.2 Os Portais

Um portal não é necessariamente corporativo. Pode ser aberto, como em muitos sítios da *Internet*. A idéia comum é sempre o conceito de ponto integrado de um conjunto de conteúdos, que podem ser expressos tanto por documentos textuais digitais como por ponto de acesso a dados armazenados em bases relacionais ou de aplicativos muito estruturados. Conforme DIAS (2001) há duas formas de classificar os portais: uma em relação ao contexto de sua utilização (público ou corporativo) e outra em relação às suas funções (suporte à decisão e/ou processamento cooperativo). Não obstante as semelhanças tecnológicas, os portais públicos e os portais corporativos atendem a grupos de usuários diversos e têm propósitos completamente diferentes. De acordo com essa autora, temos as seguintes definições:

> *Portal público: também denominado portal Internet, portal web ou portal de consumidores, provê ao consumidor uma única interface à imensa rede de servidores que compõem a Internet. Sua função é atrair, para o seu site, o público em geral que navega na Internet. Quanto maior o número de visitantes, maior a probabilidade do estabelecimento de comunidades virtuais que*

potencialmente comprarão o que os anunciantes daquele site têm para vender. Assim como a televisão, o rádio e a mídia impressa, o portal público estabelece.

Portal corporativo: no mundo institucional, o portal tem o propósito de expor e fornecer informações específicas de negócio, dentro de determinado contexto, auxiliando os usuários de sistemas informatizados corporativos a encontrar as informações de que precisam para fazer frente aos concorrentes . O portal corporativo é considerado por Reynolds & Koulopoulos[73] *como uma evolução do uso das Intranets, incorporando, a essa tecnologia, novas ferramentas que possibilitam identificação, captura, armazenamento, recuperação e distribuição de grandes quantidades de informações de múltiplas fontes, internas e externas, para os indivíduos e equipes de uma instituição .*

Os portais corporativos – ou de informações empresariais – constituem a camada superior de integração em uma organização, relacionando-se muitas vezes com EAI, ERP e outras aplicações, mas focando, porém uma comunicação dinâmica com o usuário. DE acordo com TURBAN (2003):

"um portal de informações empresariais (enterprise information portals) são aplicações baseadas na tecnologia web que permitem que as empresas acessem as informações armazenadas interna e externamente, e fornecem aos usuários um único ponto de acesso às informações personalizadas, necessárias para tomar decisões empresariais abalizadas. Os EIPs integram aplicações de gerenciamento de conteúdo, inteligência empresarial, bancos de dados e data warehouses" (p. 233).

Os portais corporativos constituem um ambiente informacional voltado para o usuário de forma a permitir uma visão sinótica (ou seja, que pode ser visto em conjunto, com um único olhar) do complexo ambiente informacional de uma organização. Ele dá uma visão integrada de aplicações e conteúdos diversos dispersos nos ambientes WEB de uma instituição, permitindo que diversos itens sejam vistos ao mesmo tempo numa mesma tela. Um portal é um padrão para acesso diverso a conteúdos e aplicações. São ambientes personalizados que viabilizam o acesso a várias informações sobre uma dada empresa. A princípio, o alvo são os funcionários, mas gradativamente pode incluir fornecedores, clientes e parceiros de uma organização. Constitui-se pois numa moldura, que permite acesso ao universo informacional da empresa, encapsulando dados e

[73] Referência apresentada pela autora: REYNOLDS, Hadley & KOULOPOULOS, Tom. Enterprise knowledge has a face. *Intelligent Enterprise*, v. 2, n. 5, p. 29-34, Mar. 1999. [online], abril 2000. http://www.intelligententerprise.com/993003/feat1.shtml

funções necessárias para que alguém realize o que deseja com a conciliação de múltiplas informações. Cria, portanto:

- um único ponto de entrada garantindo segurança de sites e aplicações;
- um único ponto de acesso ampliando a capacidade de navegação e busca, limitando e organizando privilégios de acesso e a formatação e apresentação de informações;
- um ponto único de contato para solicitação de processamento;
- um ponto único de integração.

Os portais estão se transformando cada vez mais no *front-end* da WEB, servindo como uma plataforma que podemos chamar *de auto-serviço*, acessando múltiplas fontes de informação e conteúdo. Dentre os fatores adversos relativos à interface com o usuário que pretende tratar e que fazem de sua importância algo crescente podemos citar:

- Existência de plataformas e formatos proprietários;
- Dificuldade em acessar informações vitais
- Redundância e duplicação de informação
- Dificuldade para indivíduos em publicar informação para o resto da empresa
- Vários métodos para se buscar e acessar informação
- Necessidade de Intermediação de TI para acessar informação
- Excesso de Informação
- Ausência de mecanismos de colaboração – interna e externa
- Dificuldade de integração de dados
- Aplicações de TI focada na facilidade da estação de trabalho

Cada vez mais os portais tendem a se transformar em plataformas dinâmicas, atraentes e sofisticadas para prover acesso a múltiplos conteúdos, dinamizando a criação, difusão, organização e compartilhamento do conhecimento. Suas facilidades agilizam a transformação do conhecimento tácito em conhecimento explícito.

Constituem-se em ferramentas potenciais para gerir conhecimento. Além de tornar mais dinâmico o acesso aos conteúdos de informação, podem, por isso mesmo, apoiar processos como ensino à distância, implantação de universidades corporativas ou virtuais, treinamento e desenvolvimento de funcionários. Podem também servir à ampliação da cidadania, dentro do conceito governo eletrônico, através do qual, órgãos públicos prestam serviços e fornecem informações aos seus cidadãos.

Este capítulo descreveu algumas tecnologias da informação que se articulam e apóiam o processo de gestão do conhecimento nas organizações. Não pretendeu ser exaustivo, mas apenas apresentar algumas tecnologias importantes para tal gestão, sendo que a mineração de textos, tecnologia central nessa tese, será abordada no próximo capítulo.

8 A MINERAÇÃO DE TEXTOS

8.1 Considerações Preliminares

A geração de documentos textuais está ligada à própria escrita. Podemos dizer que um dos tipos de registro mais simples que podem ser gerados é o textual. Pessoas produzem textos com facilidade e os editores de textos e páginas *web* tornaram isso ainda mais fácil. Nos últimos anos assistimos a um marcante crescimento da capacidade de processamento e armazenagem dos computadores, que tiveram sucessivas ampliações de seus espaços em discos. Esse crescimento permitiu também aos usuários um processo de crescimento da armazenagem de milhares de documentos de diversos tipos, acumulando coleções de arquivos desorganizados de difícil gerenciamento. Os usuários colecionam imagens, vídeos, fotos, novos artigos, mensagens de correio eletrônico, reportagens e notícias, etc. São resultados de uma variedade de buscas ou elaboração própria de muitos documentos texto.

Nas organizações, esse grande volume de bases de dados textuais, a ampliação do uso do ambiente WEB por pessoas e empresas, o crescimento de portais corporativos e ambientes de comércio eletrônico vem provocando uma verdadeira sobrecarga de informações, também gerando aumento do armazenamento e tornando cada vez mais complexa a procura de informações pelos diversos sítios da *Internet* e pelos repositórios das *Intranet*s. Pessoas e organizações podem saciar seus desejos de informação pesquisando documentos espalhados pelas redes sobre os mais diferentes assuntos. Entretanto, tamanho volume de documentos e informações acaba por trazer problemas na hora da pesquisa e frustração dos usuários com tal sobrecarga, que ocorre quando o usuário tem muita informação ao seu alcance, mas não tem condições de tratá-la ou de encontrar o realmente deseja ou lhe interessa.

Embora crie ilusão de que se tem muita informação disponível, a recuperação de documentos em grandes bases de textos se converte muitas vezes em frustração pela dificuldade de conhecimento do usuário sobre o real teor das bases dos documentos armazenados, pois é impossível conhecer o teor de todos eles no momento da recuperação. Os motores de busca (*search engines*) foram uma solução interessante para lidar com essa realidade, e eles se tornaram muito importantes com a difusão da *web*.

Em muitos casos, eles eram realmente uma solução poderosa para lidar com grandes bases de dados e realizar o casamento instantâneo entre uma consulta e os documentos correspondentes que a atendiam, desde que as palavras-chave utilizadas para fazer esse casamento (*matching*) estivessem corretas. Entretanto, os motores de busca não resolviam todo o problema da informação.

Os Sistemas de Recuperação da Informação tem uma longa história, marcada especialmente a partir dos anos 60 quando os sistemas computacionais eram principiantes para gerenciar textos não estruturados. Muitos desses sistemas estavam em grandes computadores denominados "*mainframes*", com softwares proprietários e interfaces para coleções especializadas. Na década de 80 os computadores pessoais fizeram o link entre esses sistemas de mainframes e interfaces intermediárias para busca e recuperação de dados. Essa ampliação, entretanto, não significou uma melhoria significativa para o usuário final. Isso porque a abordagem adotada passou a se constituir da palavra-chave ou *keyword*. Muito usada nos dias atuais, a palavra-chave pode ser definida como um substantivo ou frase encontrada no conteúdo de um documento relevante. Usuários sofisticados, com conhecimentos de lógica booleana e terminologia, podiam elaborar consultas melhores e assim encontrar as informações relevantes mais rapidamente, mas eles constituem uma exceção. A maioria dos usuários interage de forma relativamente simples com os sistemas e faz uso simplório da palavra-chave. Entretanto, tal abordagem não é suficiente para responder às necessidades mais específicas e elaboradas de informação.

Se por um lado a mineração de textos pode ser vista como uma sub-área da Recuperação da Informação, uma vez que utiliza massivamente recursos desse campo de conhecimento, por outro é também um campo interdisciplinar, considerando que é o resultado interseção das áreas de recuperação da informação, reconhecimento de padrões, banco de dados, mineração de dados, lingüística, aprendizado de máquina e estatística. Essa combinação visa permitir a célere extração de informações de grandes bases de documentos textuais. Destarte, um software de mineração de textos, quando implantado e sendo alvo de utilização pelas pessoas, constituir-se-á em um sistema de informação.

Para SULLIVAN (2001) a área de mineração de textos expande o escopo das atividades de informação estratégica (*business intelligence*). De acordo com KONCHADY (2006) a mineração de textos não é uma substituição para a Recuperação da Informação ou o Processamento da Linguagem Natural. Ela busca encontrar respostas para questões que são difíceis ou impossíveis de responder somente com motores de busca. Por exemplo:

- Sumarizar documentos que descrevem marcas de um produto consumido em uma região;

- Relacionar diferentes tipos de carros fabricados baseando-se em características comuns, tais como o tamanho, o tipo de bateria utilizada, o material da carcaça, o espaço interno, dentre outras;

- Identificar o máximo de comunidades de mergulhadores em uma região do pais e descrever algumas relações entre elas;

- Monitorar o nível de poluição por monóxido de carbono em uma cidade para o próximo mês e notificar por correio eletrônico os dias em o nível de poluição exceder o limite do tolerado;

- Identificar universidades que realizam pesquisas sobre o conhecimento em energia solar e encontrar e-mails, nomes e endereços de indivíduos para contactar nessas universidades.

Encontrar respostas para tais questões requer muito mais do que submeter algumas consultas em um motor de busca. Os resultados de uma consulta devem ser analisados, organizados e reduzidos sob uma forma de fácil entendimento. Ferramentas de mineração de textos podem usar os resultados dos motores de busca para encontrar informações escondidas, tais como tendências e relacionamentos.

8.2 – Necessidades de Textos do Usuário

Com a questão da sobrecarga de informações acima relatada, ou seja, a quantidade de documentos disponíveis nas bases de dados digitais supera em muito a capacidade das pessoas de lerem estes documentos para deles extraírem conhecimento, passamos a nos defrontar com o problema de como extrair conhecimento de grandes coleções de textos, de forma dinâmica e automática, já que os motores de busca não poderiam responder a essa necessidade. Quando um usuário, em uma organização, realiza uma busca de informações em uma fonte de dados textuais (a *web*, por exemplo) sua primeira tarefa é separar aqueles documentos que foram recuperados, mas que, claramente, não atendem às suas necessidades. A seguir, realiza uma leitura preliminar de cada um dos outros documentos, pois muitos deles podem conter informações interessantes, porém relacionadas a algum outro setor da empresa, para onde podem ser encaminhados. Ao final desta primeira leitura, o funcionário terá identificado os documentos que possuem informações relevantes e pertinentes à sua área de atuação. Passa, então a fazer uma segunda leitura, mais cuidadosa, que lhe permita classificar os documentos segundo sua relevância e tema (novo ou já conhecido da empresa). Será interessante, então, buscar documentos nas bases de dados da empresa, para serem analisados em conjunto com os novos, a fim de que possam ser identificadas correlações entre eles. Portanto, esta é uma tarefa que consome muito tempo e nem sempre o volume de documentos envolvidos permite que certas correlações entre informações sejam identificadas. Ou seja, muito conhecimento contido na coleção de documentos pode passar desapercebido. A tecnologia de mineração de textos pode trazer diversos benefícios neste campo.

8.3 – Conceito e Benefícios

KDT – Knowledge Discovery in Text – Descoberta de Conhecimento em Textos, mineração de textos, ou simplesmente *text mining*, é uma tipo de tecnologia da informação composta por todas as técnicas que permitem localizar, recuperar, organizar de forma dinâmica documentos de interesse em determinadas bases textuais, articulando seus conteúdos e os próprios documentos entre si, acelerando a análise dos mesmos, gerando resumos automáticos e ‘’extraindo” conhecimento deles e que, de outra maneira, permaneceria escondido nessas mesmas bases de dados textuais. É um sistema

inteligente de descoberta de conhecimento, que absorve os aspectos analisados no capitulo 6, com foco em bases textuais. De acordo com FELDMAN & DAGAN (1995) essa abordagem surgiu da necessidade de análises automáticas em textos, visto que a sobrecarga de informações disponíveis dificultava sua análise manual, localização e acesso. Em outro trabalho FELDMAN & SANGER (2006) afirma que a mineração de textos é um processo de conhecimento intensivo em que um usuário interage com uma coleção de documentos utilizando um conjunto de recursos de análise embutidos em uma ferramenta de software, tendo recebido uma inspiração muito forte da área de mineração de dados.

FELDMAN & SANGER (2006) avaliam que o elemento chave para a mineração de dados é a coleção de documentos, desde as mais simples até as mais amplas. Essas coleções podem ser estáticas ou dinâmicas, ou seja: as coleções estáticas não recebem novos documentos enquanto as dinâmicas recebem novos documentos a uma dada freqüência. As coleções dinâmicas tendem a ser maioria. Isso significa que um software de mineração de textos deverá re-processar suas bases quando houver novas adições em suas bases de documentos. De acordo com esses autores, é também importante considerar que outro elemento fundamental para estudos de mineração de texto é o documento. Um documento pode ser simples ou complexo, pode ser um relatório, um e-mail, um artigo, um notícia, um relato. Podem estar organizados ou não dentro de uma coleção ou dispersos de forma *ad hoc*. Podem também pertencer a diferentes coleções ou sub-coleções de documentos.

Os algoritmos utilizados em mineração de textos podem operar sobre diferentes níveis de características de representação dos documentos, devendo levar em conta aspectos semânticos e computacionais para o processamento. Em geral, as ferramentas de mineração de textos podem atuar sobre os caracteres, as palavras, os termos e frases extraídas do *corpus* dos documentos e, ainda, os conceitos (FELDMAN & SANGER, 2006).

Text mining é um conjunto de técnicas e processos que identifica conhecimento novo nos textos. Trata-se pois, até certo ponto, de uma nova tecnologia: software para análise inteligente e mineração de dados textuais. Não se trata de um navegador ou simples motor de busca. A idéia é a de que, considerando uma base de textos

(documentos digitalizados), possamos ter um software que faça leituras inteligentes desses textos, com análise semântica, agrupamento de textos comuns e inferências inteligentes, usando para isso recursos computacionais avançados ligados à área de inteligência artificial. É tecnologia de uso recente no Brasil e serve para apoiar atividades estratégicas de gestão do conhecimento, marketing estratégico, pesquisa de mercado, auditoria, gestão de bases documentais, inteligência competitiva, e complemento às funções de gerência do relacionamento com o consumidor (*CRM*). A mineração de textos e a mineração de dados tem muito em comum, variando basicamente o objeto a ser tratado: dados estruturados no primeiro caso e textos digitais no segundo. A mineração de textos absorve quase todas as visões e métodos descritos nos itens 6.3, 6.4 e 6.5.

Dentre os principais benefícios do processo de mineração de textos, podemos destacar alguns. Inicialmente, a análise de documentos recuperados em uma busca fica muito facilitada, pois os documentos já serão previamente analisados e agrupados em *clusters*[74], com a indicação de seu provável conteúdo, e serão exibidas informações sobre a relevância de cada documento, em relação ao tema considerado. Um outro ponto é que é possível a realização de buscas pró-ativas, nas quais o programa analisará os documentos sem a necessidade de uma hipótese inicial para a busca, podendo identificar novos temas relacionados aos interesses da empresa. Um terceiro aspecto é o de que os documentos poderão ser arquivados em conjunto com outros sobre o mesmo tema automaticamente, sendo ainda possível a identificação de informações correlatas nas bases de dados internas da empresa. Um quarto ponto é o de que os esforços humanos para a leitura dos documentos são menores, pois há a utilização do recurso de geração automática de resumos nesse tipo de ferramenta. Por fim, a manutenção dos dicionários temáticos pode ocorrer de maneira contínua, a partir da análise dos documentos.

Obviamente um software para mineração de textos não "irá dizer" tudo o que é importante, substituindo o especialista humano que seria encarregado da tarefa de estudo e análise inteligente de documentos. O que ele faz é tomar seu lugar no trabalho repetitivo e volumoso, valendo-se de algoritmos inteligentes que lhe permitem fazer

[74] *Cluster*: agrupamento de documentos que tem pontos em comum e se articulam em uma determinada linha. Há um alinhamento entre eles ao abordar certos assuntos e conceitos, o que pode passar desapercebido pelo usuário.

uma pré-leitura dos documentos, analisando-os de maneira preliminar. A tecnologia atua realizando inferências que resultam na separação dos dados e documentos irrelevantes, agrupamento dos semelhantes (*clusters*), identificação de correlações entre informações dispersas em documentos distintos e geração automática de resumos, a fim de que o usuário possa ter uma idéia do conteúdo da coleção, antes mesmo de ler os documentos. O especialista poderá, então, selecionar para estudo apenas aquilo que considerar relevante, com ganho precioso de tempo e produtividade, uma vez irá o texto integral apenas do que for realmente importante, aproveitado em atividades mais específicas. Interagindo com o programa, o especialista pode direcioná-lo para obter resultados cada vez mais de acordo com seus interesses, o que contribui para que atinja seus objetivos com um esforço bastante reduzido.

Um software de mineração de textos pode armazenar informações sobre os temas que interessam a diversos setores da empresa e, se for configurado para isso, realizar análise inteligente dos resultados das buscas na *web*, automaticamente (*web mining*), monitorando os documentos recuperados, para agrupá-los e encaminhá-los aos setores correspondentes, além de enviar alertas para avisar da chegada de novas informações. Aquelas informações que interessarem a mais de um usuário podem também ser armazenadas em um único local, sendo um alerta enviado a cada um desses. Evidentemente, para que o software funcione adequadamente, devem ser providenciadas a criação e manutenção de dicionários (contendo informações sobre os idiomas suportados e os temas de interesse) e devem ser estabelecidas regras criteriosamente escolhidas para a análise dos documentos (criação de clusters, geração de resumo e atribuição de grau de relevância) e para sua distribuição.

8.4 Implementação de Software de Mineração de Textos

Os algoritmos computacionais que tornam isso possível estão baseados em recursos matemáticos e estatísticos, em agentes inteligentes apoiados por técnicas de inteligência artificial e em novas e inteligentes maneiras de fazer uso do poder de processamento do computador (ZANASI, 1997). Para SULLIVAN (2001), *Text Mining* é uma disciplina nova, envolvendo diferentes áreas, cujos fundamentos são a recuperação da informação, a lingüística computacional e o processamento da linguagem natural, já descritos nessa tese no capítulo 6.

A mineração de textos implica na extração de padrões ou conhecimentos interessantes e não-triviais a partir de documentos textuais. A tecnologia de *text mining* pode ser aplicada para formalizar e explorar conhecimento tácito. O conhecimento disponível com pessoas pode ser em grande parte convertido em textos, os quais serão analisados para se entender seu significado, ou seja, do que tratam os textos. Depois, é possível explorar o conhecimento extraído dos textos para gerar novos conhecimentos. Também se pode combinar este conhecimento com o conhecimento explícito armazenado em bancos de dados estruturados. Hoje em dia, o problema é encontrar a informação útil entre os dados disponíveis. Por exemplo, Inteligência Competitiva através do *Text Mining* pode ser definida como o processo de descoberta ou predição das decisões estratégicas dos concorrentes e/ou o entendimento das características do negócio pela utilização de técnicas quantitativas de análise aplicadas em bases de dados abertas (públicas).

FELDMAN & SANGER (2006) sugerem que deve haver um relacionamento entre o usuário e o software de mineração de textos. Na medida em que há entradas de documentos e seu respectivo processamento, temos as saídas de análises, padrões, tendências e relacionamentos existentes nessa base. O usuário interage com essas saídas, avaliando os resultados, adicionando ou subtraindo elementos e mesmo sugerindo outras consultas, conforme sugere a figura 32 abaixo:

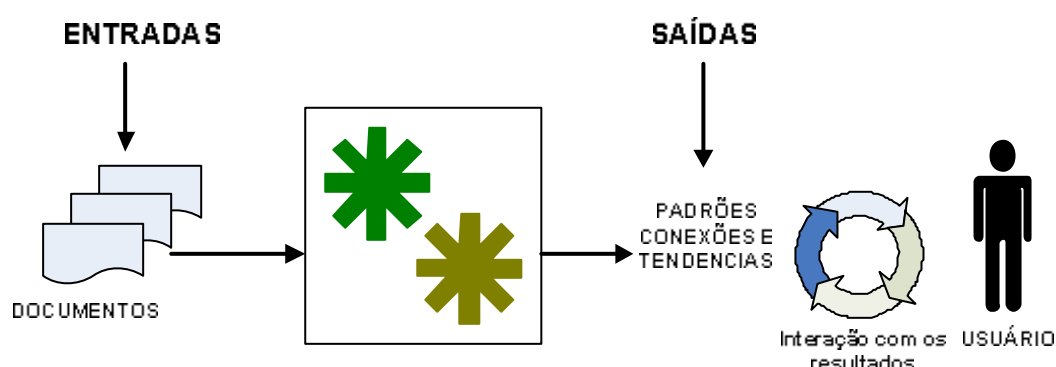



Figura 32: Relacionamento Interativo do Usuário Com Entradas e Saídas
Fonte: Adaptado de FELDMAN & SANGER (2006, p. 14)

8.5 Facilitadores e propiciadores da expansão da Mineração de Texto.

De acordo com ZANASI (2005), conforme já visto, o primeiro deflagrador é a explosão de dados, com o crescimento acelerado de dados textuais nas organizações. Muitas empresas consideram o crescimento dos dados textuais como uma interessante oportunidade para detectar novas oportunidades e realizar novos negócios. Usar esses dados corretamente é como desenvolver uma poderosa ferramenta de resultados operacionais. Um segundo deflagrador é o valor dado por empresas e governos é a busca de uma extração precisa e pertinente de informação crítica, seja para fins de segurança nacional seja visando a sobrevivência corporativa. O desenvolvimento, na última década, do campo da mineração de texto tem simplificado significativamente recuperação da informação documental para uso pessoal e profissional.

O mais importante deflagrador é o desenvolvimento tecnológico e novas pesquisas relacionadas à recuperação da informação, ao processamento da linguagem natural e tratamento semântico de textos, bem como a aprendizagem de máquina. Pesquisa e desenvolvimento recente de novas descobertas que criam novas aplicações para a tecnologia de mineração de textos nas empresas e organizações governamentais estimulam novas iniciativas. Completando os esforços e iniciativas, temos que empresas de software e universidades vem se dedicando ao desenvolvimento do campo de novas e significativas tecnologias de mineração de textos, e isso pode ser classificado em três categorias:

a) Reconhecimento inteligente de texto: a habilidade de conhecer o contexto gramatical e relacionamentos lógicos entre conceitos dentro dos textos;
b) Classificação inteligente de texto: a habilidade de classificar documentos entre um ou outro conjunto pré-definido ou de gerar categorias automáticas.
c) Manejo e processamento de múltiplos idiomas: a habilidade de trabalhar simultaneamente com documentos escritos em diferentes línguas ou que contem diferentes línguas.

ZANASI (2005) apresenta uma discussão que pode ser sintetizada na figura 33 a seguir e que compara certas questões colocadas pelos usuários em confronto com problemas relatados e exemplos a serem discutidos.

QUESTÃO	PROBLEMA RELATADO	EXEMPLO
Entender o contexto gramatical	Avaliação da satisfação do consumidor	Um e-mail que contém "Eu estou insatisfeito" será classificado como e-mail de cliente insatisfeito.
Entender relacionamentos lógicos entre conceitos	Extração e entendimento de ligações entre conceitos, pessoas ou organizações, expressas dentro dos documentos.	No campo da Inteligência Competitiva, uma empresa será constantemente informada sobre sua imagem em reputação, e aqueles competidores que tem sinalizado uma nova parceria ou anunciado o lançamento de um novo produto. No campo da Inteligência Governamental, conexões entre pessoas e organizações são exibidas em alguns segundos.
Representação de informações chave	Exploração dos resultados de uma consulta, especialmente se ela gerar mais do que 100 documentos diferentes	Representação automática de e-mails recebidos através do serviço de atendimento ao consumidor
Taxonomias: organização de documentos recuperados dentro de categorias pré-definidas.	Armazenamento automático de um documento em uma categoria pré-definida	Geração automática de relatórios inteligentes sobre todos os documentos recebidos (ex. número de e-mails contendo um mesmo conceito: observações negativas sobre um determinado produto, ou ameaças de ações terroristas)
Extração de informações simultâneas de documentos em diferentes línguas	Análise de uma coleção de documentos em diferentes línguas em uma base de dados	Análise do *feedback* dos clientes ou declarações políticas de competidores chave sobre um mesmo tópico ou assunto realizadas em diferentes línguas
Extração de identificadores pessoais chave	Identificação de autor	Detecção de características pessoais do autor que está em anonimato nos textos
Detecção de pessoas com antecedentes e experiências pessoais similares ou estilos de escrita parecidos	Detecção de redes terroristas	Detecção de ligações entre indivíduos

Figura 33: Questões solucionadas por mineração de textos.
Fonte: ZANASI (2005)

8.6 Abordagens da Mineração de Textos

Podemos reconhecer duas abordagens em mineração de textos: a primeira é a abordagem da interpretação inteligente e semântica, resultado da mesclagem das estratégias de *data mining* com análise semântica e a segunda a estatística. As abordagens estatísticas estão vinculadas à análise da importância dos termos em função da quantidade de vezes que ele aparece em um documento. As aplicações estatísticas (mostrando a distribuição dos dados e de suas varáveis) permitem que sejam detectados os elementos que vão orientar a análise. Exemplos:

- Análise de Freqüência: contando o número de referências sobre a informação procurada – os tópicos que estão sendo trabalhados (muito freqüente); - novas metodologias (bastante freqüente); - ruídos (muito raramente). Pode ser útil em algumas situações.

- Técnica de formação de pares: Realçam as conexões entre documentos.

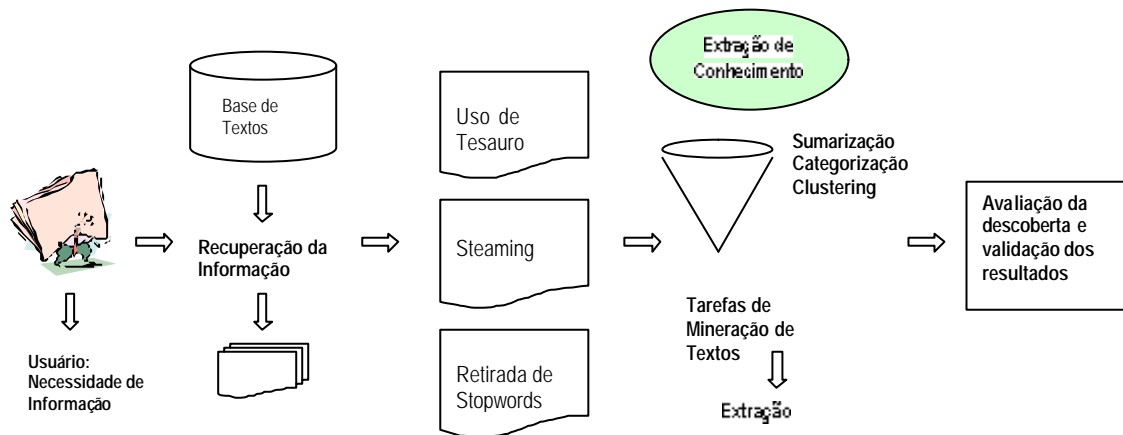
As abordagens de inteligência e semânticas de mineração de textos vinculam-se a questões como a avaliação da seqüência de termos no contexto da frase, para a correta identificação da função de cada termo, utilizando fundamentos e técnicas baseadas no processamento da linguagem natural (REZENDE, 2003), descritos no tópico 6.5 da presente Tese. REZENDE (2003) salienta o aspecto da análise da estrutura semântica dos textos:

> *"A análise semântica dos textos é feita para tentar identificar a importância das palavras dentro da estrutura da oração. Quando se utiliza um único texto, algumas funções podem ser identificadas e pela função identifica um grau de importância. Por exemplo, um nome tem grande chance de representar informação relevante na estrutura de uma sentença. Porém, para algumas tarefas esse tipo de análise não é suficiente. Para tarefas como a categorização, o interessante seria analisar um documento comparando-o com bases de conhecimento de diferentes assuntos para descobrir a qual categoria ele pertence* (p. 342)."

Essa abordagem também inclui as técnicas de mineração de dados. A mineração de textos é um processo semelhante á mineração de dados. A diferença está no objeto:

documentos X dados. O processo descrito no tópico 6.4 se aplica à mineração de textos. Algumas técnicas foram aplicadas e melhoradas nos últimos anos: redes neurais, algoritmos genéticos, associação de algoritmos, análise de grupo, arvores de decisão. Os dados de entrada são os extraídos de bases, que em seguida são tratados e analisados. As variáveis escolhidas entre as áreas de referência levantadas, serão usadas, para fins de análise, como por exemplo, descritores de documentos.

8.7 Etapas de Mineração de Textos[75]

Há diferentes abordagens possíveis para implementação de *text mining* mas que, em geral, seguem etapas mais ou menos comuns. As etapas da mineração de textos, a partir de SULLIVAN (2001) e REZENDE (2003) englobam as atividades de pré-processamento dos dados, execução das tarefas de processamento dos dados e análise dos resultados. Segue-se então uma atividade de avaliação das descobertas e validação dos resultados. Esse esquema é apresentado na figura a seguir.

coleções de documentos digitais. No pré-processamento se aplicam vários modelos descritos no capitulo 5, especialmente os modelos booleano e do espaço vetorial.

Na etapa seguinte, temos a análise dos dados, que visa facilitar a identificação de similaridade de significado entre as palavras, não obstante suas variações morfológicas. Assim sendo, o processo de mineração de textos deve inicialmente identificar e remover as *stopwords* (palavras desnecessárias ao contexto de busca e pouco relevante para análise de textos, tais como preposições, pronomes, artigos, etc.). Outro aspecto importante da análise de dados é o processo de *stemming* (identificação da raiz de palavras) com vista a apoiar uma análise semântica[76]. Também para análise dos dados é possível utilizar um tesauro (dicionário) como um vocabulário controlado que representa sinônimos, hierarquias e relacionamentos associativos entre termos para apoiar usuários na descoberta de informação que eles necessitam. Ele ajuda a solucionar certos problemas da linguagem natural, uma vez que usuários diferentes podem usar termos diferentes para uma consulta equivalente. Por isso, um dicionário controlado define termos variantes (sinônimos, abreviações, acrônimos, siglas, escritas alternativas, etc.) para um termo preferido e único para cada conceito.

Em seguida temos, a etapa de processamento dos dados, em que é realizado o processo de indexação e realizadas as tarefas de descoberta de conhecimento em textos. A indexação permite a procura eficiente em textos por documentos relevantes a uma consulta sem a necessidade de exame dos documentos por inteiro. Os tipos de indexação mais comuns, conforme REZENDE (2003), são:

- Por texto completo – ocorre automaticamente em várias ferramentas de mineração de textos quando os documentos são carregados.
- Temática – realizada com apoio de um dicionário controlado (tesauro) que fornece uma estrutura hierárquica que permite ao software de mineração de textos encontrar rapidamente generalizações e especializações de termos específicos
- Indexação Semântica Latente – já descrita no capitulo 5.

[76] Há diferentes algoritmos para realizar essa tarefa, tais como os métodos de Stemmer, de Porter de Lovins (REZENDE, 2003, p. 348).

- Indexação por tags – algumas partes do documento são selecionadas para fazer parte do índice.

As tarefas de descoberta de conhecimento são apresentadas na seção 8.7.

A última etapa da mineração de textos é o pós-processamento ou avaliação de resultados, que pode ser feita com base em métricas[77], processos de visualização e analise qualitativa dos resultados por especialistas.

8.8 Formas de Descoberta do Conhecimento em Mineração de Textos

As tarefas de mineração de dados descritas no tópico 6.4.3 se aplicam também à mineração de textos, podendo ser vistas como formas de descoberta de conhecimento em textos. Há uma convergência entre diferentes autores do campo da mineração de textos em descrever essas tarefas ou formas de descoberta (SULLIVAN, 2001; WIVES, 2002; REZENDE, 2003; ZANASI, 2005; KONCHADY, 2006; FELDMAN & SANGER, 2006), quais sejam: sumarização, categorização, *clusterização* (*clustering*), extração de características ou de informações. Por se tratar de um campo emergente da pesquisa, não há ainda um consenso quanto à melhor forma de se classificar suas possíveis abordagens. Múltiplos procedimentos são possíveis e diferentes abordagens também, as quais estão sendo propostas por diferentes autores. Podemos citar alguns procedimentos utilizados, em conjunto ou isoladamente.

- Extração de características: procura extrair de um texto ou conjunto de textos somente informações mais relevantes para o usuário, uma vez que se não fossem extraídas provavelmente passariam despercebidas. Estão vinculadas ao processamento da linguagem natural.

- Clusterização ou agrupamento: agrupar em classes elementos que possuem alguma articulação ou elementos similares. Inicialmente cria-se uma representação simplificada de cada texto a ser adicionado aos agrupamentos. Em seguida, determina-se a proximidade de dois documentos com base em seus vetores de características. Os documentos são agrupados com base em alguma

[77] Índices de precisão e revocação, por exemplo.

medida de similaridade (REZENDE, 2003). É um método de descoberta de conhecimento utilizado para identificar correlações entre e associações entre objetos, facilitando assim a identificação de classes. Esse processo é usado geralmente usado anteriormente à classificação, ajudando a definir classes, pois um usuário especialista é capaz de analisar correlações entre os elementos de uma coleção de documentos e identificar uma melhor distribuição de classes para os objetos em questão. Os clusters podem ser isolados ou hierárquicos. No primeiro caso, criam-se grupos distintos de documentos agregados. No segundo caso, cria-se uma estrutura de relacionamento ou ligação entre os grupos, o que pode permitir um processo de navegação, como em uma árvore (WIVES, 2002).

- Categorização: implica em estabelecer classes e definir a qual classe ou categoria um documento pertence. Visa identificar características principais de um documento e inseri-lo em categorias pré-definidas. Isso é feito com base em análise de modelos lingüísticos, como afinidades léxicas e freqüência de palavras. Conforme WIVES (2002), pode também usar regras de inferência, modelos conexionistas (redes neurais artificiais) ou similaridade de vetores ou de centróides, em que as classes são representadas por vetores (conjuntos) de palavras (denominados centróides).

- Sumarização: reduzir a quantidade de texto em um documento sem perda de seus significados mais importantes. Identifica palavras, frases e muitas vezes conceitos mais importantes de um documento para permitir uma visão geral e preliminar de um documento ou conjuntos de documentos, permitindo ao usuário maior agilidade em seu trabalho. Extrair sentenças e palavras-chave ajuda o processo (ZANASI, 2005). A extração de características e técnica do centróide são úteis a esse processo.

- Análise de trechos: descobrir conhecimento por resgate de trechos permite aos usuários encontrar detalhes de informação, sem que necessitar ler todo texto. Entretanto, ainda assim, é preciso que o usuário leia e interprete as partes do texto que forem recuperadas para extrair a informação desejada.

- Análise lingüística e de conteúdo: informações e regras podem ser descobertas através de análises lingüísticas em nível léxico, morfológico, sintático e semântico;

- Associação de trechos: busca encontrar automaticamente conhecimento e informações relacionadas no mesmo texto ou em textos diferentes;

- Uso de estruturas de textos: segundo MORRIS (1991), determinar a estrutura de um texto ajuda a entender seu significado. Um texto não é um conjunto aleatório de frases, mas deve haver uma unidade e também coesão, com as frases funcionando juntas para a função do todo. Um *thesaurus* é utilizado para determinar as relações que são significativas entre termos próximos, sendo analisadas associações diretas, indiretas (transitivas só em primeiro grau) e categorias comuns de termos.

8.9 Funcionalidades de um Software *Text Mining*

Um software de *Text Mining* visa oferecer suporte informatizado ao usuário para recuperação, tratamento e análise de informações textuais, através um processo de análise inteligente de conjuntos de documentos, possuindo três objetivos básicos:

- Propor um modelo que possa ser utilizado no tratamento de documentos textuais especializados como dados semânticos disponíveis em um sistema de informação, que permitam gerar informações aos profissionais de uma área especializada sem a necessidade de convertê-los em dados estruturados tradicionais.

- Propor os elementos necessários para a definição e manipulação de bases de dados semi-estruturadas que possam ser utilizados em áreas especializadas do conhecimento, tais como, medicina e outras, onde seja necessário algum tipo de classificação taxonômica para recuperação de informação.

- Servir de base para a definição e implementação de um sistema que possua a capacidade de auxiliar os usuários (especialistas e não especialistas) na navegação

sobre documentos relevantes, permitindo uma interação ampla na construção de consultas necessárias à geração da informação.

Podemos colocar como grandes requisitos de um software de mineração de textos, com base em ZANASI (2005):

- Primeiro, o aplicativo deve ser capaz de realizar buscas reativas em bases de dados textuais, recuperando arquivos a partir de palavras ou expressões de entrada, ordenando os resultados com base na contagem de palavras, organizando esses arquivos em clusters e fornecendo o resumo dos arquivos solicitados.

- Segundo, o aplicativo deve ser capaz de realizar buscas pró-ativas em subconjuntos específicos dessa base de dados, identificando quais são as idéias centrais de um documento, quais delas são comuns a todo o conjunto e quais são particulares a um ou mais documentos e, se possível, encontrar relações entre estas idéias na forma de *"a idéia A normalmente aparece junto com a idéia B"*.

Exemplos de funcionalidades presentes em *softwares* de mineração de textos podem ser:

a – Gestão de Bases de Dados Textuais – Text Warehousing

Gerencia a organização e o acesso aos dados da base de forma organizada. Visa manter a base de dados atualizada sem a necessidade de intervenção constante por parte do usuário. A gerencia dessas bases textuais é amplamente discutida por SULLIVAN (2001).

b – Pré-Processamento dos Textos

Prepara documentos para serem processados por outras funções (descritas a seguir). Automatiza o processo de adequação dos documentos, liberando o usuário para tarefas mais especializadas e adequadas ao seu contexto de trabalho.

c – Geração Automática de Resumo

Diminui a quantidade de texto a ser lido pelo usuário, reduzindo o tempo de análise de um documento.

d – Criação de Clusters

Promove o agrupamento de documentos afins, permitindo a visualização de relações recíprocas entre os documentos e sua análise numa perspectiva sinóptica, o que não seria facilmente perceptível numa postura de análise “ad hoc” dos documentos. A análise de clusters é uma função poderosa que permite ao usuário extrapolar seus horizontes de análise, identificando possibilidades e perspectivas não imaginadas anteriormente, analisando, conceitos, tendências, comportamentos, relações entre diferentes fatos e realidades que sejam relevantes em seu contexto de trabalho.

e – Busca Pró-ativa

Descobre novidades e aspectos relevantes em textos, agilizando a percepção de tendências e o comportamento de um fenômeno ao longo do tempo, apresentando ao usuário informações tempestivas e significativas.

f – Armazenamento

Mantém em arquivos cópias integras dos textos analisados, tendo como benefício relevante o arquivamento, na base de documentos local, dos resultados obtidos com a análise e acrescentando informações que facilitem a recuperação dos documentos.

g – Busca Reativa

Recuperação de informações, promovida diretamente pelo usuário, a partir de critérios de busca. Tem como benefício a recuperação dos documentos de forma

organizada, agrupando-os e classificando-os de acordo com critérios informados pelos usuários.

h – Dicionários e Parâmetros

Contempla a necessidade de definição de conhecimento das palavras e dos temas relevantes. Oferece suporte para a criação e manutenção de dicionários de idiomas e temáticos.

i – Regras Para Análise e Distribuição

Gerencia dados de parâmetros para a definição das funções de resumo, cluster, armazenamento e relatório. Oferece suporte à criação e manutenção de regras para geração automática de resumo, agrupamento de documentos, classificação em termos de relevância e distribuição dos mesmos em diferentes setores da organização.

j – Categorização Automática

Possibilidade de direcionar documentos a classes previamente estabelecidas por usuários.

k – Geração de Relatório

Informa aos usuários sobre a chegada de novos documentos nas bases. O usuário recebe informações sobre os novos documentos de forma organizada, facilitando sua seleção para uma análise mais aprofundada.

8.10 Mineração de Textos e Inteligência

A informação relevante está, cada vez mais, articulada a uma perspectiva de inteligência. O conceito de inteligência está inflacionado em nossa cultura possuindo múltiplas interpretações. Portanto, requer aqui uma definição que pode ser expressa num ponto de vista organizacional. Sob este ponto de vista, inteligência vincula-se à

definição da postura estratégica da organização, ao pensar executivo sobre o conjunto articulado de ações que deve uma corporação realizar para atuar em um ambiente dado, considerando seu público alvo, *players*, produtos e intentos de participação em seu mercado de atuação. Ou seja: considerando a missão organizacional, a função de inteligência questiona a realidade, procura interpretá-la, compreender o jogo de forças sob o qual está submetido uma corporação, as fragilidades e potencialidades de todos os entes envolvidos e, uma vez entendida a configuração desse jogo e os eventos contextualizados, define o comportamento adequado da organização. Podemos, portanto, falar de uma inteligência organizacional que considera o desafio da corporação de decidir posturas estratégicas e definir ações sobre a realidade que tornem efetivas a sua missão. Essa inteligência deve ser capaz de prospectar o ambiente, captando sinais que indiquem qual é o seu estado num dado momento e imediatamente disparando comandos que alterem o comportamento corporativo para se ajustar a condições que mudam continuamente.

A mineração de textos visa a exploração de informação em documentos e é uma atividade estratégica que nasceu do desafio de acessar de modo “inteligente” a informação existente em acervos nas organizações e bases públicas na web . Por “inteligente” queremos dizer que os sistemas que possibilitam esse acesso devem ter capacidade para auxiliar o usuário que está analisando informação, interagindo com ele, mostrando-lhe os conteúdos relacionados com aquele(s) que ele busca, permitindo-lhe expandir de modo controlado o seu pedido de informação e reduzir, em seguida, rápida e interativamente, o leque de documentos encontrados e relevantes. Durante a busca e a análise, gradativamente pode sugerir ainda outros documentos relacionados com aqueles que este usuário está procurando, embora não contendo nenhuma das palavras ou expressões anteriormente utilizadas, permitindo também novas seleções desses documentos.

Uma ferramenta de mineração de textos permite a exploração de grandes quantidades de texto para daí extraírem automaticamente diversos tipos de conhecimento que esses textos encerram. A exploração de informação que realiza visa no essencial extrair conteúdos e descriminar os significados das palavras (e unidades textuais mais complexas) recorrendo aos contextos sintáticos em que elas ocorrem. Pode recorrer também a uma análise estatística dos conjuntos de documentos e dos

termos que neles figuram. A avaliação da qualidade da informação extraída será outro aspecto que será enfatizado, tendo em vista as aplicações a que aquela informação se destine. Mineração de textos é um esforço significativo de decodificação automática de segredos guardados pelas coleções de textos, segredos que muitas vezes pode significar maiores lucros ou melhores resultados pela identificação em tempo hábil de oportunidades e ameaças, pela evidência de fatos relevantes em processos internos e aprimoramentos dos controles organizacionais e pela adequada análise do mercado e das relações com os clientes.

8.11 Aplicações de Softwares de Mineração de Textos

Há um conjunto de possibilidades para aplicação do processo da Mineração de Textos, envolvendo inteligência organizacional, inteligência governamental, relacionamento com pessoas, gestão pública, desenvolvimento científico e tecnológico, gestão do conhecimento e gestão de operações organizacionais. Em todos esses campos, temos grandes bases de dados textuais, a maioria de crescimento rápido pelo acelerado desenvolvimento da tecnologia da informação e pela facilidade da geração da informação textual nos dias atuais. Podemos avaliar sumariamente alguns desses campos.

Um primeiro campo de aplicação é a Inteligência Competitiva, que pode ser definido como um programa ético e sistemático para capturar, reunir, analisar e gerenciar informação externa que podem afetar as organizações, seus planos, decisões e operações. (SCIP – *Society of Competitive Intelligence Professionals*). Sob o ponto de vista da Inteligência Competitiva, é possível aplicar as estratégias de mineração de textos com objetivo de entender as variáveis relativas aos consumidores, aos competidores, fornecedores, desenvolvimento tecnológico e desenvolvimento de novos produtos e serviços.

Um segundo campo de aplicação é a inteligência governamental vinculada à gestão pública. Governos são cada vez mais demandados em termos de eficiência e competência em suas ações pelos cidadãos de seus países e regiões, especialmente os democraticamente eleitos. Precisam responder com celeridade e efetividade no campo

das tarefas que lhes são atribuídas constitucionalmente. Temos então uma gama de aplicações para Mineração de Textos, envolvendo a melhoria da inteligência governamental, no sentido de atuar em tarefas como: detecção de fraudes e atentados contra o patrimônio e a receita públicos, gerenciamento e coordenação de ações de segurança nacional, previsão de ações terroristas, aplicação de regras de direito público, dentre outras. As organizações públicas lidam com gigantescos acervos de textos e as atividades de regulação e monitoramento conduzidas pelo Estado podem muito se beneficiar de *Text Mining*.

Um terceiro ponto importante a ser citado é o relacionamento com pessoas, inseridas em diferentes contextos e comunidades cujo monitoramento inteligente seja justificável. Temos então duas frentes importantes: o CRM – *Customer Relationship Management* (Gerencia do Relacionamento com o Consumidor), o CZRM – *Citizens Relationship Management* (Gerência do Relacionamento com os Cidadãos). Tanto enquanto consumidores de bens e serviços como enquanto cidadãos de um determinado estado, as atitudes, pensamentos e comportamentos das pessoas interessam a empresas e governo. A Mineração de Textos pode ser usada para capturar, portanto, seus comportamentos, reações, sentimentos, através da extração de informações de e-mails, *chats*, fóruns de discussão, comunidades virtuais, etc.

Um quarto campo de aplicação muito importante é a Gestão do Conhecimento. A experiência operacional de muitas organizações é frequentemente traduzida em documentos textuais, produzidos por profissionais de diferentes matizes, grande parte deles com elevada competência e formação. Documentos em formato de editores de texto, planilhas eletrônicas e páginas da *Web* são produzidos de modo intenso, diariamente. Com o passar do tempo, a base de documentos textuais cresce e é necessário saber como explorá-la para finalidades estratégicas, sob pena de, por maior que ela seja, nada servir à organização que a gerou. Muitas pesquisas para responder perguntas do presente podem basear-se em questões já tratadas no passado. A reconfiguração e nova combinação de assuntos e, consequentemente de documentos anteriormente armazenados diversos ajuda a responder desafios do presente. Bases de dados externas adquiridas ou simplesmente acessadas pela organização podem também compor o universo de interesse. Enfim, temos uma coleção de textos, a que podemos

chamar de *document warehouse*, produzida de diferentes formas pela organização que será então alvo do processo de mineração de textos.

Se a *web* e outros repositórios internos são as fontes ou "minas" de onde se extrai material bruto, *documenent warehouses* são as usinas e refinarias que tornam as pesquisas com textos usáveis em um processo de gestão do conhecimento e *business intelligence* e distribuição do mesmo ao consumidor final. *Document warehouses* são repositórios dinâmicos, constantemente adaptados diante das necessidades de mudanças apresentadas pelos usuários finais de informação. SULLIVAN, 2001

8.12 O Software ***DifBrain®***

Entre 2004 e 2006, foi desenvolvido o *software* ***DifBrain®***, pela Diferencial Consultores Associados®, [78] empresa de Tecnologia da Informação, especializada em desenvolvimento de sistemas de informações, com foco em softwares gerenciais, e sediada em Belo Horizonte, desde 1996. O *software* ***DifBrain®*** é aplicativo destinado a realizar o processo de mineração de textos. Ele realiza todas as funcionalidades descritas no item 8.9. Esse desenvolvimento foi conduzido por um grupo de trabalho composto por três profissionais, incluindo o autor da tese, que trabalhou em conjunto com um analista de sistemas e uma professora doutora da área de engenharia elétrica[79] com atuação na área de inteligência artificial.

DifBrain® foi desenvolvido de acordo com o Processo Interativo de Desenvolvimento de Software e a Orientação a Objetos, apresentado no tópico 3.5, dessa Tese, em ambiente SUN® Java® – J2EE® incorporando também recursos em PROLOG[80] para

---

[78] Maiores informações em www.difnet.com.br. Desde de 1996 a empresa atua com o desenvolvimento de softwares gerenciais, especialmente os que fazem uso de recursos tecnológicos da área de Inteligência Artificial em estratégias de "mineração de dados". Atuou também em projetos de gestão do capital intelectual de organizações entre 2000 e 2004 e outros projetos diversos vinculados aos campos da tecnologia da informação, gestão de talentos humanos e mapeamento e análise de processos.

[79] Além do autor, compuseram e equipe de desenvolvimento o Analista de Sistemas Marcelo Gibson de Castro Gonçalves, Cientista da Computação e a Professora Doutora Regina Fátima Gibson Gonçalves, Engenheira Eletricista, especialista na área de Inteligência Artificial.

[80] Linguagem de programação PROLOG – PROgramming in LOGic – ou programação em lógica. Essa linguagem é amplamente usada em aplicações da área de Inteligência Artificial. Ela é usada principalmente como uma linguagem de prototipação rápida e para tarefas de manipualação de símbolos e

as rotinas que demandaram recursos de inteligência artificial. A ferramenta passou por uma série de versões, tendo iniciado seus primeiros testes com a avaliação de noticias sobre política de dois jornais brasileiros. Posteriormente foi aplicada à base de documentos da experiência operacional interna da ELETRONUCLEAR. Incorpora de forma mesclada e integrada recursos descritos para recuperação da informação (capitulo 5) e sistemas inteligentes (capitulo 6), dentre os quais podemos citar:

- modelo vetorial;
- modelo de indexação semântica latente;
- modelo das listas não sobrepostas
- modelo da proximidade de nós
- modelo da recuperação por passagens
- modelo fuzzy
- modelo contextual
- processamento da linguagem natural.

O software em questão, utilizando modelos de RI e IA procede o tratamento do texto "quebrando-o" em palavras, apurando freqüência das mesmas e ordenando-as por número de ocorrências. Realiza a separação por sentenças e executa o tratamento das mesmas Identifica também a estrutura e segmentos do texto. Antes de seu uso na ELETRONUCLEAR, foi utilizado em testes com noticias extraídas da Internet de jornais brasileiros com sítios na *Web*. Essa ferramenta foi desenvolvida para tratar documentos em diferentes idiomas, sendo que hoje já incorporou a análise de documentos em português, inglês e alemão. Possui também as funcionalidades de buscas proativas e reativas e procede o controle de *clusters*, procedendo cruzamento entre critérios de elaboração. Cria *clusters* hierárquicos de modo a permitir a visualização por parte do usuário dos conjuntos de documentos e de seus relacionamentos criando uma perspectiva sinóptica de análise. Apresenta informações estatísticas sobre o documento. Na geração do resumo, permite que o usuário especifique o número de palavras que deseja no resumo, até o total de 200 palavras. Possui interface interativa e amigável e auxilia o usuário a realizar buscas inteligentes e céleres.

---

análise de linguagem natural. Muitos sistemas especialistas foram escritos em PROLOG para domínios jurídicos, médicos, financeiros e outros (RUSSEL & NORVIG, 2004).

PARTE III – DESENVOLVIMENTO DA PESQUISA

9 O CONTEXTO DA ENERGIA NUCLEAR E A ELETRONUCLEAR

O presente capítulo tem por finalidade apenas apresentar uma visão panorâmica sobre a energia nuclear para que se possa perceber a dimensão do conhecimento humano envolvido no âmbito da Eletronuclear, nos processos de projeto, construção, operação e manutenção de usinas nucleares e, por conseguinte, a importância dos conhecimentos tácitos, que estão embutidos nas mentes dos funcionários da empresa, e explícitos, que estão inseridos em suas bases documentais e de dados. Por conseqüência, infere-se uma visão preliminar sobre os desafios relativos à gestão desse conhecimento em um contexto organizacional. Em hipótese alguma pretende ser uma visão exaustiva de assunto tão vasto e complexo. A Engenharia Nuclear é um campo de conhecimento específico e a operação de usinas tem uma dimensão muito ampla que não é intenção da presente tese aprofundar. Procuramos apenas tangenciar o assunto, pontuando sua relevância para se justifique ainda mais as práticas de gestão do conhecimento e a tecnologia de mineração de textos a ela vinculada. Para uma visão mais ampla e profunda dos conceitos e processos relativos à energia nuclear consultar na *Internet*, por exemplo, o sítio www.nucleartourist.com, cujas informações também contribuíram para a visão aqui apresentada.

9.1 Histórico da Energia Nuclear

Um marco importante na história do conhecimento humano é a descoberta do átomo, que é fruto do desenvolvimento intelectual do homem na sua relação com a natureza e trouxe muitos benefícios ao homem em diferentes campos da ciência. Um passo adicional foi dado ao se descobrir que se poderia usufruir das características desse átomo na geração de energia. As usinas nucleares de operação comercial produzem energia elétrica a partir da energia nuclear. Por energia nuclear entendemos a força que mantêm os componentes do núcleo de um átomo integrados. Portanto, quando há o romper desta força produz-se energia térmica. Em um reator, portanto, o objetivo é romper esta força e usar a enorme quantidade de energia térmica liberada para gerar vapor, que produzirá movimento e por fim, eletricidade. Assim, a energia elétrica, em usinas nucleares, não é gerada diretamente dos átomos que se dividem.

A partir da evolução do conhecimento da estrutura atômica constatou-se a possibilidade do núcleo de um átomo poder perder massa. E verificando-se então uma perda de massa, tal perda irá ser transformada em energia. No ano de 1938, os alemães Hahn e Strassmann conseguiram fissionar o urânio. Quando um núcleo pesado sofre fissão, obtêm-se átomos de massa mediana e grande quantidade de energia. Posteriormente novos experimentos constataram a "quebra" do núcleo do urânio por meio de nêutrons. Em tal quebra, muitos produtos da fissão podem ser gerados, ou seja, temos uma sucessão de reações nucleares ocorrendo ao mesmo tempo. Nas quebras ocorrem liberações de 2 ou 3 nêutrons que, como elementos desencadeadores da fissão provocam outras sucessivas cisões nucleares (reações em cadeia). Há dois tipos de reações nucleares: a fissão e a fusão.

Na fissão, o núcleo de algum elemento químico pesado (urânio, plutônio ou tório) recebe o bombardeio de um nêutron e subdivide-se em duas partes maiores e alguns nêutrons, sendo então alvo de uma cisão que é chamada de fissão nuclear (Figura 35 a). Nela, a soma das massas resultantes tem massa inferior à soma das massas do núcleo bombardeado e do nêutron.

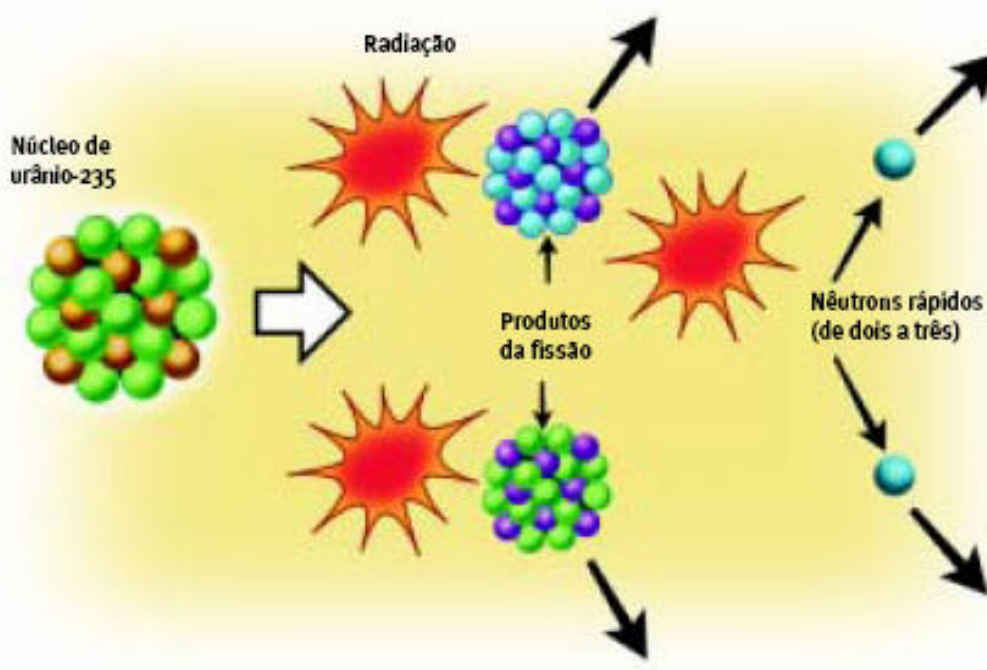



Figura 35 (a): Fissão Nuclear

O outro tipo de reação nuclear é denominado fusão. Nela, juntam dois núcleos atômicos para formar um, o que requer uma quantidade muito elevada de energia para vencer a repulsão eletromagnética inerente entre estes núcleos. O núcleo de um elemento químico leve [isótopos do H_2 (deutério e trítio)] recebe um nêutron ou dois núcleos de elementos leves [isótopos do H_2 (deutério e trítio), He e Li], que se fundem, produzindo, em ambos os casos, um elemento químico mais pesado, mas as partes resultantes da fusão têm massa inferior à soma dos elementos iniciais. As reações por fissão nuclear presidem os processos de geração comercial de energia elétrica atualmente no mundo, através da tecnologia dos reatores de fissão.

9.2 O Combustível Nuclear

Um elemento combustível (urânio, tório ou plutônio) é necessário à operação de usinas nucleares. Esse elemento combustível, em geral o urânio, precisa passar por um processo de enriquecimento, uma vez que o urânio natural não é usado para gerar energia. É necessário usar urânio-235 (U-235) que é altamente fissionável, misturado ao urânio-238 (U-238) é bem mais estável, o que permite controlar a energia produzida na reação em cadeia através da proporção entre eles. Usinas nucleares usam uma proporção de 3% de U-235 e 97% de U-238. Os reatores nucleares de Angra I e II utilizam como elemento combustível pastilhas sólidas, contendo dois tipos de átomos de urânio. Um deles, o Urânio U-235, constitui menos de 1% do urânio natural, mas fissiona facilmente. O outro tipo, o Urânio U-238, constitui os outros 99%, sendo, entretanto, praticamente não-fissionável. O urânio natural é então enriquecido, de modo que a concentração de U-235 seja aumentada para 2 a 5%. Isto perfaz uma quantidade suficiente de urânio físsil para operar satisfatoriamente uma usina nuclear.

No caso brasileiro, a opção pelo uso da energia nuclear torna-se ainda mais atraente quando se considera que o Brasil possui a 6ª maior reserva mundial de urânio e instalações industriais do Ciclo do Combustível, operadas pela INB – Indústrias Nucleares Brasileiras, o que garante ao país independência no suprimento de combustível nuclear. Para ir da mina até a usina, o urânio brasileiro percorre as seguintes etapas:

a) Mineração: prospecção, extração e o beneficiamento do minério de urânio até a produção do concentrado ou "torta de urânio", também conhecido como *yellow cake* (U_3O_8).

b) Conversão: o *yellow cake* é transformado no gás UF_6 - hexafluoreto de urânio, forma química apropriada para utilização no processo seguinte.

c) Enriquecimento Isotópico: - O UF6 passa por um processo em que se torna radioativo, uma vez que átomos do U-238 são retirados do urânio natural, obtendo-se um produto final com maior concentração de isótopo físsil, ou seja, enriquecido em U-235.

d) Reconversão: O UF6 é reduzido a óxido de urânio (UO_2), na forma de pó.

e) Fabricação de pastilhas: O pó de óxido de urânio (UO_2) é aquecido e compactado em pastilhas de 10 mm de diâmetro que têm a forma de um cilindro de cerca de 1 cm de comprimento e de diâmetro, acomodadas no interior de varetas de 4,40 m de comprimento e 10,76 mm de diâmetro de zircaloy (uma liga de zircônio), hermeticamente fechadas e que suportam temperaturas até 1.852 ºC. Para produzir uma pastilha é necessário processar 250 g de yellowcake.

f) Montagem do Elemento Combustível: As pastilhas são colocadas no interior de varetas metálicas tubulares. As varetas estanques são soldadas e montadas em grades estruturais, formando o elemento combustível. Cada vareta possui aproximadamente 2kg de urânio. Uma vareta garante fornecimento de energia a uma cidade de 20.000 habitantes, durante 24 horas. São necessárias 235 dessas varetas para formar um conjunto, o elemento combustível.

g) Geração de Energia: Um determinado número de elementos combustíveis é colocado no interior do reator. Neles ocorre a reação em cadeia, resultando na produção de calor. Em meio ao combustível, composto de pastilhas de dióxido de urânio, são inseridas pastilhas de califórnio, elemento químico que emite nêutrons naturalmente e que irá deflagrar a reação de fissão. Na fissão, um

nêutron atinge o átomo de U235 e usualmente divide-o em dois átomos menores. Esta divisão libera não apenas calor, mas também dois ou três nêutrons, que, por sua vez, atingem e dividem outros átomos. Assim, ocorre uma reação em cadeia de fissão nuclear (Figura 35 b). Após a sua "queima", o combustível pode ser armazenado em um depósito especial ou então reprocessado.

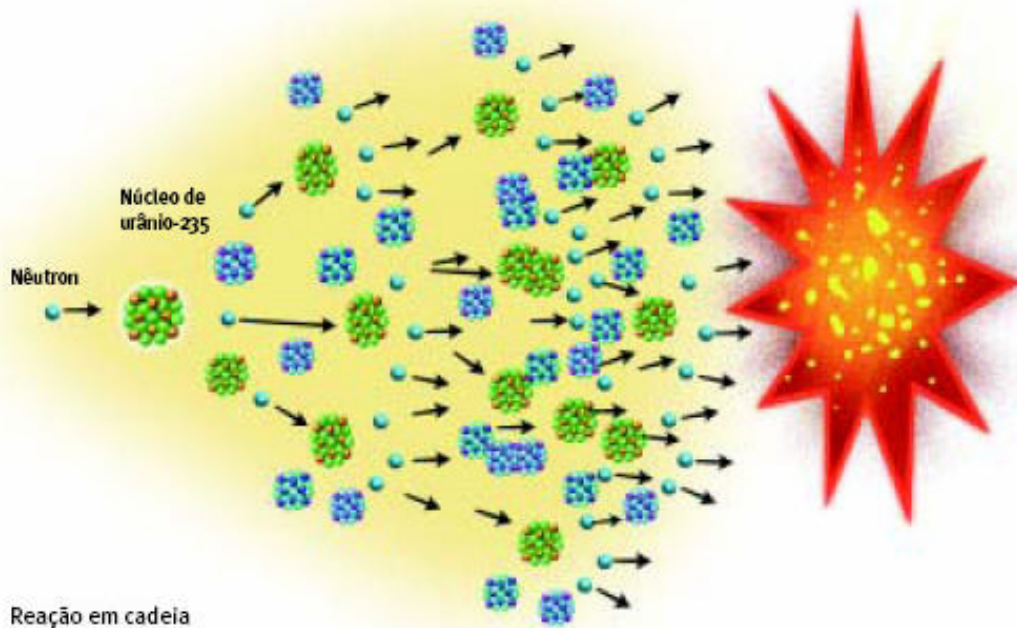



Figura 35 (b): A reação em Cadeia

9.3 A Operação de Usinas e Reatores

As Usinas Nucleares são centrais de geração de energia elétrica que utilizam a reação nuclear como fonte para geração de energia. As usinas são estruturas complexas por lidar com a geração de energia, o manejo de questões ambientais e rejeitos nucleares, além de demandar um nível elevado de conhecimento humano e recursos financeiros e materiais para operá-las. Em geral são complexos que lidam com estruturas de contenção de concreto e aço, reatores, geradores de vapor, piscinas de armazenamento de combustível, turbinas à vapor, geradores de eletricidade, condensadores, torres de refrigeração, torres de transmissão de energia, dentre outros elementos.

O reator nuclear é um elemento chave. Um reator é uma máquina térmica que gera energia a partir do calor. Um elemento combustível (urânio, tório ou plutônio) é ativado para produzir calor, que produz vapor e que por sua vez produz energia elétrica. Ativar é deflagrar a reação de fissão nuclear, que é a "quebra" dos átomos de U235 através do choque de nêutrons nestes átomos de urânio.

O núcleo de um reator consiste de um conjunto de vários tubos longos com pastilhas de dióxido de urânio, substância que contém átomos de urânio. No urânio ocorre uma reação em cadeia causada pelas fissões do urânio-235, e a energia liberada é absorvida pelo material do reator na forma de calor - a energia nuclear contida nos núcleos atômicos é transformada em energia térmica. O combustível produz o calor pela fissão e necessita de um refrigerante (água, água pesada, gás, hélio, etc). O refrigerante conduz o calor produzido durante o processo até a turbina geradora de eletricidade ou ao propulsor.

Nos reatores de água pressurizada, por exemplo, o calor é retirado do reator de modo a aquecer a água e transformá-la no vapor, que acionará a turbina. Ao passar pelas pás de uma turbina, e girá-la, a energia térmica é transformada em energia mecânica de rotação. O eixo da turbina vincula-se com um gerador que transforma a energia mecânica em energia elétrica. A energia elétrica é então conduzida, através de torres de transmissão e malhas de distribuição até residências e outros consumidores.

Além do combustível e do refrigerante, merecem menção alguns outros componentes importantes, tais como o moderador que serve para diminuir a velocidade dos nêutrons que intervêm na reação nuclear, tornando-os lentos a fim de aumentar a probabilidade de ocorrência de fissão. A moderação pode ser feita pela injeção de boro no circuito primário ou pelo uso de barras de controle. Outro item significativo é o refletor, que impede o escapamento de nêutrons aumentando a eficiência do reator. Há ainda a blindagem, os materiais de controle e os elementos de segurança.

Uma blindagem de concreto, chumbo ou aço evita o escapamento de radiação gama e neutrons. Cádmio e boro são usados na forma de barras ou dissolvidos no refrigerante como elementos de controle para finalizar uma reação em cadeia

Atualmente existem vários tipos de reatores nucleares de fissão, conforme descritos a seguir:

- LWR - *Light Water Reactors*: Utilizam como refrigerante e moderador a água e como combustível o urânio enriquecido. Os mais utilizados são os BWR (*Boiling Water Reactor* ou Reator de água em ebulição) e os PWR (*Pressure Water Reactor* ou Reatores de água pressurizada), estes últimos considerados atualmente como padrão. Havia, em 2001, 345 dessses reatores no mundo em funcionamento.
- CANDU - *Canada Deuterium Uranium*: Utilizam como moderador água pesada (cuja molécula é composta por dois átomos de deutério e um átomo de oxigênio) usando como refrigerante água comum. Como combustível usam urânio comum. Em 2001 havia 34 deles em operação no mundo.
- FBR - *Fast Breeder Reactors*: Utilizam nêutrons velozes no lugar de térmicos para o processo da fissão. Como combustível utilizam plutônio e como refrigerante sódio líquido. Este reator não necessita de moderador. Havia apenas 4 deles em operação em 2001.
- HTGR - *High Temperature Gas-cooled Reactor*: Usa uma mistura de tório e urânio como combustível, utilizando hélio como refrigerante e grafite como moderador. Existiam 34 em funcionamento em 2001.
- RBMK - *Reactor Bolshoy Moshchnosty Kanalny*: O objetivo principal desse tipo de reator é a produção de plutônio, gerando eletricidade como subproduto. O grafite é usado como moderador, a água como refrigerante e o urânio enriquecido como combustível. Pode recarregar-se durante o funcionamento. Existiam 14 em funcionamento em 2001.
- ADS - *Accelerator Driven System*: Utiliza uma massa de tório, sendo a fissão produzida pela introdução de nêutrons no reator de partículas através de um acelerador de particulas. Encontra-se ainda em fase de experimentação, e uma de suas funções fundamentais será a eliminação de resíduos nucleares produzidos em outros reatores de fissão.

Portanto, uma forma de classificar as usinas nucleares é em função do fluido que resfria o reator. Há reatores refrigerados a gás, reatores refrigerados a água fervente, reatores refrigerados a água leve moderado a grafite, reatores refrigerados a água

pressurizada, reatores refrigerados a água pesada. Há também processos operacionais com reatores a água leve fervente, reatores regeneradores refrigerado a metal líquido, reatores regeneradores rápidos refrigerados a gás e outros.

9.4 A Geração de Energia Através do Reator à Água Pressurizada (PWR)

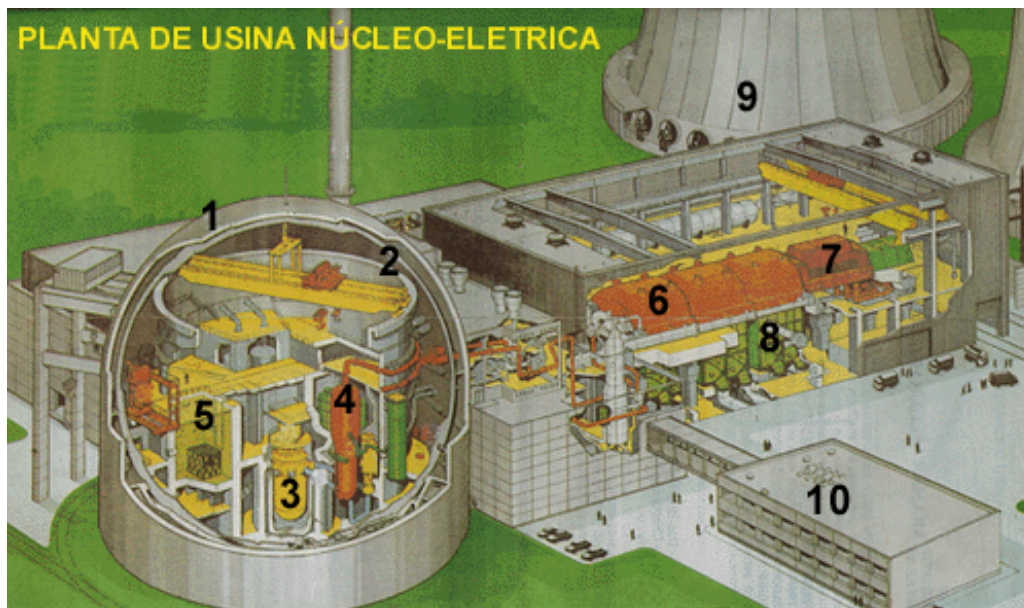
O princípio de funcionamento das usinas PWR - *Pressurized Water Reactor* (Reator a Água Pressurizada) se baseia no resfriamento do núcleo do reator através de um circuito fechado de água da alta pressão chamado circuito primário. Nelas, a água aquecida sob alta pressão do circuito primário passa por um trocador de calor (gerador de vapor) onde aquece e transforma em vapor a água do circuito secundário. Esse vapor movimenta uma turbina que aciona um gerador elétrico. A condensação do vapor que trabalha na turbina se faz num trocador de calor (condensador) que é resfriado por outro circuito dotado de uma torre de refrigeração. A energia gerada chega aos consumidores finais através de redes de distribuição. A figura abaixo mostra planta típica de uma usina nuclear:

1. CONTENÇÃO DE CONCRETO ARMADO
2. CONTENÇÃO DE AÇO
3. REATOR NUCLEAR
4. GERADOR DE VAPOR
5. PISCINA DE ARMAZENAMENTO DE COMBUSTÍVEL
6. TURBINAS À VAPOR
7. GERADOR DE ELETRICIDADE
8. CONDENSADORES
9. TORRE DE REFRIGERAÇÃO
10. ADMINISTRAÇÃO

Figura 36: Planta de Usina Núcleo-Elétrica
Fonte: CTMSP

Explicando melhor, em usinas com reatores de Água Pressurizada (PWR) é utilizado o chamado ciclo indireto: a água líquida que se encontra sob pressão no tonel do reator circula num circuito (o circuito primário) distinto do circuito onde circula a água que depois de vaporizada atravessa a turbina (circuito secundário). A formação de vapor de água no circuito secundário é obtida no gerador de vapor através da troca do calor com a água do circuito primário. O processo de geração de energia elétrica a partir da energia nuclear, então, pode ser esquematizado em três passos:

1 - No reator: transformação da energia nuclear em energia térmica, através da reação nuclear em cadeia;

2 - Na turbina: transformação da energia térmica em energia mecânica, através da ação do vapor d'água aquecido;

3 - No gerador: transformação da energia mecânica em energia elétrica.

Na figura 37 a seguir, temos outro detalhamento:

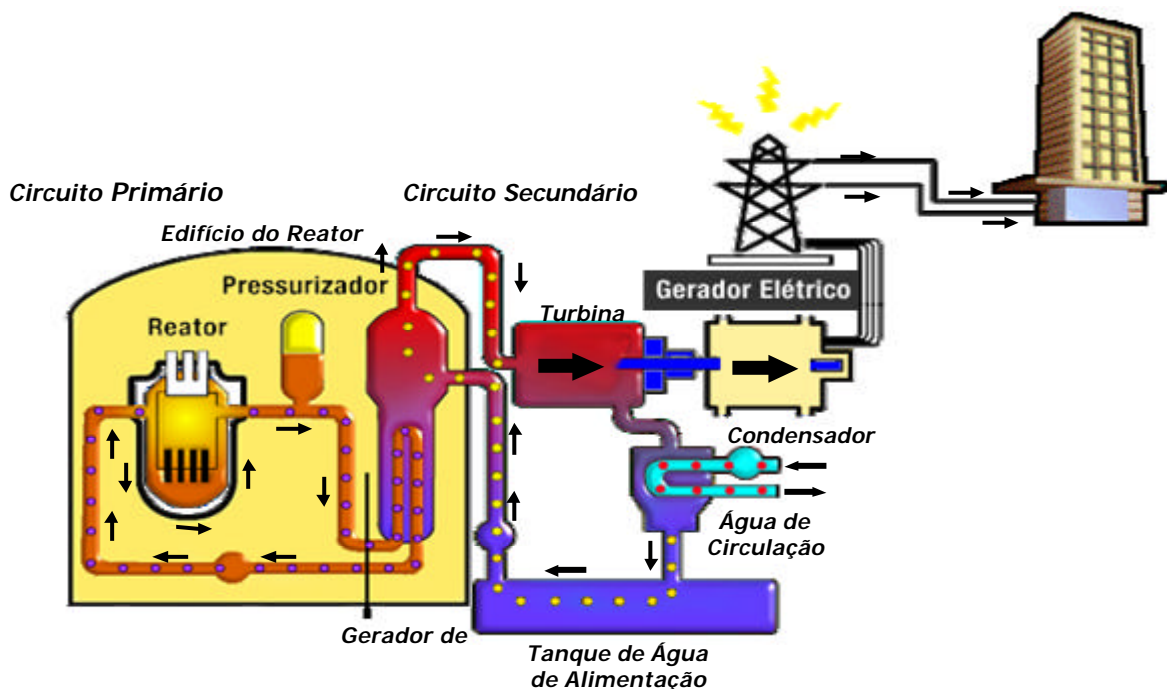



Figura 37: Geração de energia elétrica a partir de um reator PWR
Fonte: Eletronuclear

Diante do acima exposto, podemos inferir que as organizações responsáveis por planejar, implantar e gerir plantas industriais voltadas para a geração de energia são organizações que lidam com considerável base de informações sobre suas operações e tecnologias empregadas, demandando conhecimentos específicos e estratégicos cujo desenvolvimento não é simples e precisa, portanto, ser objeto de uma política de geração, manutenção e utilização.

9.5. Uso e Perspectivas da Energia Nuclear

De acordo com a WNA – *World Nuclear Association* (Associação Nuclear Mundial), em dados apresentados na *Revista Carta Capital* de 13.09.2006, podemos afirmar que o mundo já tem hoje uma significativa dependência da energia nuclear, que, conforme citado anteriormente, representa 16% da matriz energética mundial, com geração de 2,626 trilhões de KWh, e consumo de 65.478 toneladas anuais de urânio. Apresentamos a seguir o quadro com dados relevantes da WNA.

<table>
<tr><th colspan="4">DEPENDENCIA ATÖMICA</th></tr>
<tr><th colspan="2">10 maiores produtores (bilhões de KWh)</th><th colspan="2">10 maiores consumidores (em % da matriz energética)</th></tr>
<tr><td>EUA</td><td>780,5</td><td>França</td><td>79</td></tr>
<tr><td>França</td><td>430,9</td><td>Lituânia</td><td>70</td></tr>
<tr><td>Japão</td><td>280,7</td><td>Bélgica</td><td>56</td></tr>
<tr><td>Alemanha</td><td>154,6</td><td>Eslováquia</td><td>56</td></tr>
<tr><td>Coréia do Sul</td><td>139,3</td><td>Ucrânia</td><td>49</td></tr>
<tr><td>Rússia</td><td>137,3</td><td>Coréia do Sul</td><td>45</td></tr>
<tr><td>Canadá</td><td>86,8</td><td>Suécia</td><td>45</td></tr>
<tr><td>Ucrânia</td><td>83,3</td><td>Bulgária</td><td>44</td></tr>
<tr><td>Inglaterra</td><td>75,2</td><td>Armênia</td><td>43</td></tr>
<tr><td>Suécia</td><td>69,5</td><td>Eslovênia</td><td>42</td></tr>
<tr><th colspan="4">PANORAMA DA ENERGIA NUCLEAR NO MUNDO</th></tr>
<tr><td colspan="2">Geração de Eletricidade</td><td colspan="2">2,626 trilhões de KWh</td></tr>
<tr><td colspan="2">Participação na Matriz Energética</td><td colspan="2">16%</td></tr>
<tr><td colspan="2">Consumo anual de urânio</td><td colspan="2">65.478 toneladas</td></tr>
<tr><td colspan="2" rowspan="4">Reatores</td><td>Em funcionamento</td><td>442</td></tr>
<tr><td>Em construção</td><td>28</td></tr>
<tr><td>Planejados</td><td>52</td></tr>
<tr><td>Propostos</td><td>152</td></tr>
<tr><td colspan="4">Fonte: WNA – Word Nuclear Association (Associação Nuclear Mundial)
Publicado na Revista Carta Capital, São Paulo, (410) pág 16, 13.09.2006</td></tr>
</table>

Fonte: SIQUEIRA, André. Potência Nuclear? – Revista Carta Capital, São Paulo (410) set 2006

Considerando questões geológicas, ecológicas e econômicas, produzir petróleo e outros combustíveis fósseis não será mais possível no mesmo ritmo do consumo mundial de energia. Como o carvão, o petróleo e o gás natural, a energia nuclear é uma das formas de obtenção de eletricidade em larga escala. Entretanto, nas próximas décadas, os combustíveis fósseis tendem a se exaurir, ou se tornar excessivamente danosos ao meio ambiente. Observa-se que a oposição à utilização de energia nuclear

está começando a diminuir. Na Suécia, já existe um grupo ambientalista que defende o uso de energia nuclear, denominado *"Environmentalists for Nuclear Energy"* [81]. Devido à pressão de ambientalistas contra a utilização de combustíveis fósseis, assiste-se a uma crescente consideração da utilização da energia nuclear como uma opção energética plausível. Esta transformação é nítida em países de grande influência no cenário mundial, como os Estados Unidos, o Japão e a França. Esta tendência, muito provavelmente se manifestará também no Brasil.

Já temos hoje países em que a dependência da energia nuclear é muito alta, com essa fonte representando um percentual alto em suas matrizes energéticas. De acordo com a WNA – Word Nuclear Association (Associação Nuclear Mundial) na França esse percentual chega a 79%, na Lituânia 70%, na Bélgica 56% e na Suécia 45%. Há fatores macroeconômicos que hoje favorecem o uso da energia nuclear. O crescimento da economia mundial demanda mais energia de fontes cada vez mais escassas. Por conseqüência, temos aumento dos preços de varias *commodities* energéticas. As fontes fósseis estão cada vez mais caras. Os sucessivos aumentos no preço do barril indicam uma perspectiva de aumento sustentado do preço do petróleo. A crescente demanda mundial pressiona os preços do gás natural. Temos também a diminuição da possibilidade da geração de energia elétrica por recursos hídricos, uma vez que a maior parte da malha hídrica do país, próxima aos principais centros consumidores encontra-se próxima do esgotamento, além do fato de termos hoje dificuldade de aprovação de obras para novas hidrelétricas diante do alto custo de enquadramento ambiental. Há também dificuldades na exploração em grande escala de fontes alternativas como a energia solar e a eólica. Esses aspectos apontam no futuro para um aumento do uso mundial da energia nuclear, tornando remotas as possibilidades de recuo em programas de geração dessa energia no Brasil e no mundo. Isso implica em um aprofundamento do processo de busca de excelência em gestão e operação de usinas nucleares, fazendo da prática de gestão do conhecimento um recurso poderoso para o melhor funcionamento das mesmas.

---

[81] Fonte: www.eletronuclear.gov.br

9.6 Descrição da Eletronuclear [82]

No ano de 1968, o Estado brasileiro tomou a decisão de ingressar no campo da produção de energia nuclear, visando permitir ao país a possibilidade de desenvolver conhecimento com tal tecnologia e adquirir experiência para fazer frente às possíveis necessidades futuras. Assim sendo, em 1969, realizou-se uma concorrência internacional, vencida pela empresa norte-americana Westinghouse e, em 1972, iniciou-se a construção da Usina de Angra I, com uso de reator do tipo Reator a Água Pressurizada – PWR (*Presssurized Water Reactor*). A usina entrou em operação comercial em 1985.

Em junho de 1975, o Governo do Brasil assinou com a então República Federal da Alemanha o Acordo de Cooperação para Uso Pacífico da Energia Nuclear. No âmbito desse acordo, foi concretizada, em julho de 1975, a aquisição das usinas de Angra II e III à empresa alemã Kraftwerk Union A.G. – KWU, então subsidiária da Siemens. A Usina de Angra II, após atrasos de cronograma devido a restrições de recursos financeiros disponíveis, entrou em operação em 2000, sendo sincronizada à rede em julho desse mesmo ano.

Atualmente, de acordo com a WNA – *Word Nuclear Association* (Associação Nuclear Mundial), a energia nuclear responde por 16% da produção de energia elétrica total no mundo. Devido ao seu baixo custo ambiental, evita a emissão anula de aproximadamente 2 bilhões de toneladas de CO_2. Em 2001, Angra I e II evitaram a emissão de 14 milhões de toneladas de CO_2 que seriam produzidos caso a energia produzida fosse gerada pela queima de óleo combustível ou carvão, como estava previsto nos anos 60 para o Rio de Janeiro, que receberia complementação termelétrica naquela época.

A ELETRONUCLEAR é uma empresa de controle estatal e foi criada em 1º de agosto de 1997, proveniente da fusão da área nuclear de FURNAS Centrais Elétricas S.A., responsável pela operação de Angra 1 e pela construção de Angra 2, com a

---

[82] As informações a seguir sobre a empresa foram obtidas em seu endereço eletrônico da Eletronuclear (www.eletronuclear.gov.br) e em documento de apresentação da Central Nuclear Almirante Álvaro Alberto fornecido pelo Centro de Informações mantido pela Empresa em Angra dos Reis – RJ. Inclui também os relatos apresentados ao autor por ocasião das diversas visitas realizadas na Eletronuclear.

NUCLEN, empresa de engenharia detentora da tecnologia de projeto de Angra 2 e 3. Sua missão é produzir energia elétrica de origem nuclear com elevados padrões de segurança e eficiência e custos competitivos, preservando a capacidade de projetar, construir e gerenciar seus empreendimentos. A empresa possui competência e capacidade técnica nas áreas de projeto, construção e operação de usinas termonucleares. As atividades da Empresa compreendem hoje a operação da Usina Nuclear Angra 1, com 657 Mw e Angra 2, com 1309 Mw. Tais usinas, em conjunto com Angra 3, também com 1309 Mw e que no momento passa por uma fase de reavaliação do projeto, constituem a Central Nuclear Almirante Álvaro Alberto – CNAAA, situada na praia de Itaorna, no Município de Angra dos Reis, Estado do Rio de Janeiro.

Em 2001, Angra II apresentou um desempenho significativo, gerando 10,5 milhões de MWh, o que a colocou em 16º. Lugar no ranking mundial das usinas nucleares com maior volume de geração de energia. Melhorias tecnológicas introduzidas em vários sistemas e componentes conduziram a um aumento do valor nominal da potência para 1350 MW disponíveis para operação em regime contínuo, valor este posteriormente homologado, em 2002, pela agência reguladora do setor elétrico, ANEEL – Agência Nacional de Energia Elétrica.

A Central Nuclear de Angra está próxima dos principais centros consumidores de energia do país. A escolha do local envolveu uma série de condicionantes, todos ligados às características do sistema de geração nuclear, tais como abundância de água de refrigeração e facilidade de transporte e montagem de equipamentos pesados, sendo precedida de inúmeros estudos desenvolvidos com o apoio de empresas de consultoria internacionais, com ampla experiência em seleção de sítios para a construção de usinas nucleares. A proximidade dos grandes centros de consumo evita a construção de dispendiosos sistemas de linhas de transmissão e a conseqüente elevação do custo da energia produzida.

Está prevista a Construção da Usina Nuclear de Angra III, que será praticamente uma réplica de Angra II, com exceção de suas fundações, assentadas diretamente sobre a rocha, e da modernização do sistema de instrumentalização e controle, que será inteiramente digital. Os componentes dos circuitos primários e secundário de Angra III,

fornecidos nos anos 80, juntamente com os de Angra II, estão armazenados sob rigorosas condições de preservação que asseguram seu perfeito estado, mas com custos com relevantes, o que deve ser um fator importante para a rápida retomada da implantação da Usina. Os procedimentos para o licenciamento nuclear e ambiental de Angra III estão em andamento e sua operação está prevista para 2013, de acordo com o Plano Decenal da Empresa de Planejamento Energético (EPE) do Ministério das Minas e Energia. A usina está incluída entre as fontes de energia brasileiras previstas para o país, portanto. A conclusão da Usina Nuclear Angra III representará uma valiosa contribuição para o abastecimento de energia do Estado do Rio de Janeiro e da Região Sudeste. Neste novo mercado, altamente competitivo e que é a principal razão da reformulação do setor elétrico, é meta primordial da Empresa minimizar o custo de produção da energia elétrica de origem nuclear, mantendo os mais elevados padrões de segurança no projeto, construção e operação das usinas que integram a CNAAA. A retomada da construção de Angra III implica em um futuro crescimento do uso da energia nuclear no Brasil, que hoje responde por cerca de 4,3% do total da energia consumida no país.

As questões de segurança são relevantes na operação de usinas nucleares e obedecem a especificações internacionais. O licenciamento das usinas da CNAAA está a cargo da CNEN – Comissão Nacional de Energia Nuclear, obedecendo ainda de forma rigorosa à legislação ambiental vigente no Brasil. As usinas são também avaliadas periodicamente por organismos internacionais como a IAEA (*International Atomic Energy Agency*), WANO (*Word Association of Nuclear Operators*) e INPO (*Institute of Nuclear Power Operators*).

Outra questão importante e sempre considerada na operação de usinas nucleares é o tratamento de rejeitos. Os rejeitos líquidos e sólidos, de média e baixa atividade, gerados durante a operação das usinas da CNAAA são processados e acondicionados em tambores testados e qualificados segundo as normas da CNEN e requisitos internacionais. Após devidamente lacrados e identificados, esses tambores são armazenados no Depósito Provisório de Rejeitos, especialmente construído para esse fim e localizado na área central. Os tambores estão sob fiscalização permanente e são periodicamente inspecionados e monitorados. Em 2001 a Eletronuclear e a CNEN firmaram acordo para construção de um Depósito Final de Rejeitos de baixa e média

atividade gerados pelas usinas. Esse local ainda está em fase de escolha, sendo o término previsto para sua construção em 2009.

São objetivos estratégicos da Eletronuclear:

- Otimizar a produção de Angra 1 e Angra 2, apresentando padrões de segurança e desempenho acima da média;
- Implementar um programa de gerenciamento de rejeitos, com foco na redução dos volumes produzidos no armazenamento em caráter intermediário e na obtenção de solução para o armazenamento definitivo;
- Retomar a construção de Angra III;
- Atuar, ativamente, na condução do Programa Nuclear Brasileiro e na divulgação dos benefícios da energia nuclear para o país;
- Dotar a empresa de instrumentos que permitam, continuamente, a realização profissional de seus empregados, o desenvolvimento de seus processos corporativos, o seu equilíbrio econômico-financeiro e o gerenciamento do seu conhecimento.

A Eletronuclear, sendo uma empresa que atua na área da energia nuclear, tem como um de seus objetivos principais dominar a respectiva tecnologia, conforme OLIVEIRA (2) (2002)[83]. Ainda segundo esse autor, Tal tecnologia foi absorvida e consolidada ao longo dos últimos 30 anos, pelas empresas que passaram a constituir a ELETRONUCLEAR: a NUCLEN e a Diretoria Nuclear de FURNAS. Após 1997, pela

---

[83] O Sr. Luiz Celso Oliveira, funcionário da Gerência de Engenharia de Apoio da Eletronuclear, sendo *Coordenador de Experiência Operacional Externa - EOE*, é autor do trabalho *Gestão do Conhecimento na Eletronuclear* e foi o principal interlocutor da empresa com o autor da presente tese no que tange ao levantamento das informações imprescindíveis para sua realização, bem como facilitador do processo de interação com demais pessoas e áreas técnicas e operacionais da organização. Atuou também no fornecimento de dados e informações que apoiaram a especificação do software de mineração de textos *DifBrain®*, usado nos experimentos realizados na Eletronuclear com suas bases de documentos digitalizados, sendo, juntamente com sua equipe de trabalho, um dos principais usuários potenciais de tal tecnologia dentro da Empresa. O trabalho citado foi gentilmente apresentado ao autor dessa tese, muito contribuindo para a melhor compreensão da realidade organizacional a ser pesquisada e facilitando a busca de soluções para problemas apresentados, dentro do escopo da metodologia de pesquisa do presente trabalho, que implica em uma construção coletiva de propostas e alternativas para desafios que se apresentam na situação em análise.

própria ELETRONUCLEAR. A NUCLEN, como parte do Acordo Nuclear Brasil – Alemanha, assimilou a tecnologia para projeto, construção e gerenciamento do empreendimento de usinas nucleares tipo *PWR*, através de um processo planejado e monitorado. FURNAS absorveu da *Siemens* e da *Westinghouse*, através de treinamento e licenciamento de operadores, os conhecimentos necessários para operar o mesmo tipo de usina. Uma condição de fixação da tecnologia é a sua prática constante, mas com o enorme espaçamento entre a construção das usinas nucleares brasileiras (duas usinas em 30 anos, no lugar de nove usinas em 20 anos, originalmente planejadas pela ELETROBRÁS), a manutenção da tecnologia ficou ameaçada. Isto principalmente pela não renovação suficiente das equipes técnicas, o que não era requerido pelo ritmo lento da construção das usinas. operação e de apoio de engenharia com o melhor domínio possível das respectivas tecnologias. Ainda de acordo com OLIVEIRA (2) (2002), da análise da estratégia da Eletronuclear, e de seu mercado de atuação, podem ser os seguintes os seus fatores críticos de sucesso:

- Baixo custo de operação
- Alta disponibilidade das Usinas
- Atingir os indicadores de performance da WANO (*Word Association of Nuclear Operators*)
- Performance química
- Exposição coletiva a radiação
- Razão de segurança industrial
- Volume de resíduos sólidos
- Desempenho dos sistemas de segurança
- Performance Térmica
- Fator de perda de disponibilidade não planejada

Pelos aspectos descritos, podemos perceber que a questão da gestão do conhecimento na Eletronuclear se apresenta com status de objetivo estratégico, devido a peculiaridade e relevância de sua missão organizacional, objetivos estratégicos e fatores críticos de sucesso, enfim para sua sustentabilidade no curso da história e dentro do mercado de energia nuclear.

10 INFORMAÇÃO E CONHECIMENTO NA ELETRONUCLEAR

10.1 Histórico

A partir da descrição do capítulo anterior, podemos perceber que a gestão de usinas nucleares deve ser entendida com um sistema sociotécnico complexo. Uma estrutura operacional de geração de energia a partir de fonte nuclear, demandando um conjunto amplo de equipamentos, materiais e serviços é ponto de partida. A operação de usinas nucleares é uma atividade estratégica e que deve ser conduzida por profissionais sérios, competentes, éticos e capazes de alocar um volume significativo de conhecimento. Por sua vez, esse conhecimento possui especificidade e complexidade tal que dificulta a disponibilidade de grandes contingentes de mão-de-obra qualificada. Além disso, há variáveis estratégicas e de segurança, relacionadas ao processo de geração energia nuclear que ampliam ainda mais o aporte de conhecimento necessário à gestão das usinas. Portanto, baseando-se na análise de OLIVEIRA (2) (2002), que é um marco importante para o processo de gestão do conhecimento da ELETRONUCLEAR, podemos dizer que o uso da tecnologia nuclear demanda uma grande amplitude de conhecimento concernente a projeto, construção, operação e manutenção de usinas nucleares. Tal demanda envolve produção, acesso, uso e recuperação de informações associadas a esse conhecimento e que estão embutidas em pesquisa científica, estudos e análises das diversas especialidades da engenharia, documentação de projeto, informações de operação, registros de manutenção, inspeções regulatórias e outros diversos documentos relativos à informação técnica. Tudo isso é mesclado e imbricado na cognição e na *expertise* das pessoas que trabalham com tais conhecimentos e que são capazes de fazer uso adequado desses objetos de informação, deles produzindo novos conhecimentos a cada dia e/ou aplicando-os nos processos de trabalho. Cientistas, engenheiros, gestores, profissionais de diversas disciplinas com formação educacional e a perícia requeridas devem ser capazes de aplicar esse conjunto de conhecimentos de modo seguro e eficiente. Ainda de acordo com OLIVEIRA (2) (2002):

> *"O efetivo gerenciamento do conhecimento nuclear inclui assegurar a contínua disponibilidade deste reservatório essencial de pessoal qualificado. O gerenciamento do conhecimento nuclear é crítico para garantir a certeza de segurança, encorajando a inovação, criando a certeza que os benefícios da energia nuclear relacionados com a saúde humana, alimentos e agricultura,*

tratamento da água, suprimento de energia elétrica, e uma gama de outras aplicações, permaneçam disponíveis para gerações futuras" (p. 2).

Em 2000, a Resolução da Conferência Geral (44)/RES/21 da AIEA – Agência Internacional de Energia Atômica valorizou as práticas de gestão do conhecimento em usinas nucleares. Salientou a importância de itens como implantação de plano de sucessão para pessoal detentor do conhecimento, preservação da memória institucional e do conhecimento operacional sobre plantas nucleares. A partir de tais práticas, reconheceu-se já por tal Agência que nos Estados Unidos as usinas puderam aumentar a vida útil, sua potência e melhora da imagem. Conforme a referida resolução, IAEA (2002):

> *"o desenvolvimento e o gerenciamento seguro da geração nucleoelétrica, e da ciência e suas aplicações, nos países membros que decidiram ou irão decidir pela opção nuclear, requer a preservação do conhecimento e da experiência, bem como a manutenção de potencial de recursos humanos tecnológica e cientificamente competentes na esfera nuclear."*

RINTALA & KURONEN (2006) afirmam que as organizações nucleares estão lutando para preservar sua experiência e conhecimento apesar da difundida previsão de aposentadoria de muitos profissionais capacitados em um futuro próximo. As operações da plantas nucleares estão sentindo, manifestando e noticiando que a aposentadoria de profissionais experientes tem levado à perda do conhecimento nuclear tácito, devido à dificuldade de codificar e/ou compartilha-lo durante o treinamento formal. O compartilhamento de conhecimento tácito, especialmente com a amplitude e complexidade dos relacionados à operação de plantas nucleares, demandam métodos baseados em interação social, considerando situação por situação ou cada grupo envolvido. Portanto, é necessário adotar métodos de compartilhamento de conhecimento tácito para preservar o conhecimento nuclear ao longo do tempo.

Para RINTALA & KURONEN (2006), há uma preocupação generalizada entre as organizações nucleares, agentes reguladores e empresas operadoras de plantas nucleares, no que tange à preservação do conhecimento nuclear. Hoje há menos estudantes de ciência nuclear, bem como declinou a pesquisa e desenvolvimento na área. Além disso, temos muitas aposentadorias, sendo que os que se aposentam possuem

tanto o conhecimento explicito quanto o conhecimento tácito, acumulado em muitos anos de trabalho a partir de situações, experiências e vivencias de execução de tarefas.

Em 2001, além de colocar a gestão do conhecimento como pré-requisito para a retomada da construção da usina nuclear de Angra 3, a Eletronuclear constatou risco de perda de conhecimento devido ao índice elevado de funcionários com idade superior a 45 anos (Figura 38), estando muitos deles próximos à aposentadoria. Isso está associado a uma tendência, em nível mundial, em que a força de trabalho está envelhecendo, ou seja, mais e mais trabalhadores da área nuclear estão se aproximando da idade de aposentadoria, sem o correspondente influxo de pessoal jovem adequadamente qualificado para substituí-los. A questão que passou a se apresentar, portanto, é a de como gerenciar o conhecimento estratégico da empresa, identificando seus componentes tácitos e explícitos, avaliando as possibilidades de explicitação através do uso de tecnologias da informação de forma a manter o patrimônio intelectual da organização. Ao comentar, em recente entrevista, sobre os equipamentos já adquiridos pelo Governo Brasileiro e a necessidade do Brasil chegar a uma conclusão sobre Angra III, o presidente da Associação Brasileira de Energia Nuclear (Aben), Edson Kuramoto, alertou para o risco crescente de a bagagem tecnológica acumulada pelo País se perder com o tempo:

> *"Não é só o uso dos equipamentos que está em jogo. A média de idade dos especialistas é de 50 anos e dentro de mais dez anos a maioria deles vai se aposentar. Pelo menos 200 engenheiros brasileiros foram enviados para a Alemanha, nas décadas de 70 e 80, para receber treinamento"* [84]

[84] SIQUEIRA, André. Potência Nuclear? In Revista Carta Capital, São Paulo (410) 10-13, set. 2006

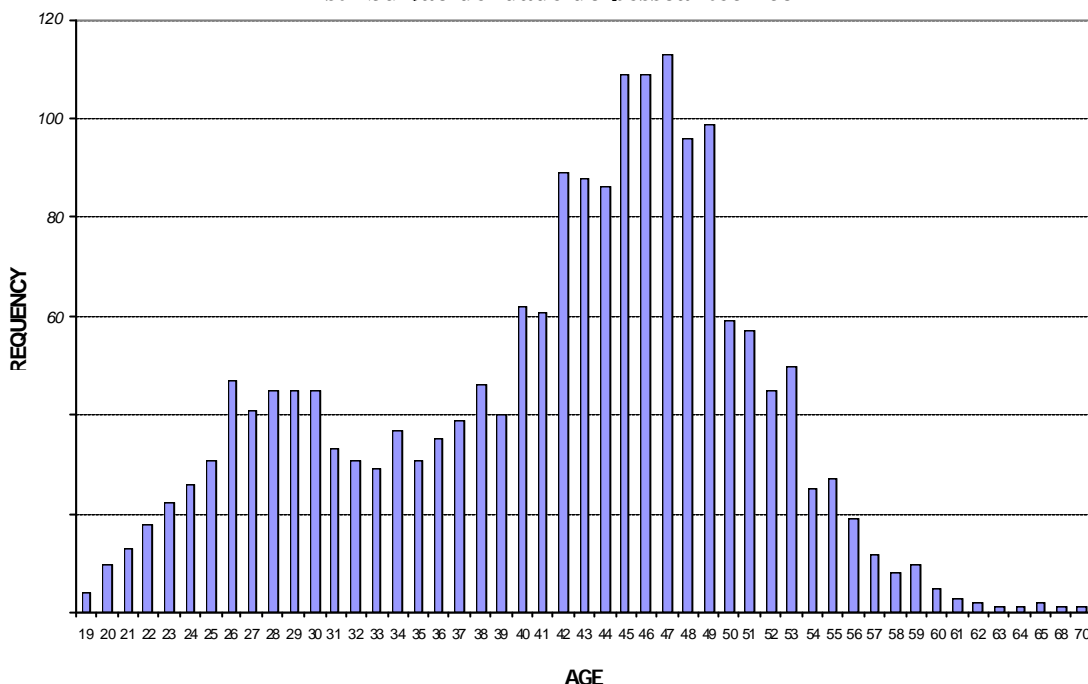



Figura 38: Distribuição dos Empregados por Idade na ELETRONUCLEAR (áreas técnicas)
Fonte: LEPECKI, 2002

OLIVEIRA (2) (2002) colocou como dois grandes desafios para a ELETRONUCLEAR a necessidade de um Plano de Sucessão e a preservação da memória organizacional, ao afirmar:

> *"Uma questão fundamental está relacionada ao "Plano de Sucessão" para a força de trabalho da área nuclear, assegurar que quando cientistas, engenheiros e técnicos se aposentarem, uma nova geração com formação apropriada e aspirações de carreira esteja disponível para ocupar seus lugares. Renovar a força de trabalho é uma questão com implicações potenciais em segurança e confiabilidade das instalações. Competência técnica para a operação segura das instalações nucleares já existentes, bem como para salvaguarda de material nuclear, pesquisa e desenvolvimento, gerenciamento e transporte de resíduos e descomissionamento de instalações, serão necessárias por décadas, indiferentemente da ocorrência ou não de expansão no uso da energia nuclear. Outra questão importante envolve a potencial perda da "memória institucional" com a saída de pessoal. Baseado em sua experiência de trabalho estas pessoas podem possuir conhecimento de fatos e procedimentos essenciais não documentados que podem ser perdidos. Isto passa a ser extremamente importante com o envelhecimento das instalações e quando são necessárias modificações de projeto. Isto exige ações gerenciais para "preservação do conhecimento" (explicitar e transferir conhecimentos*

antes da saída de pessoal). Para um encaminhamento adequado das questões apresentadas, a Gestão do Conhecimento pode contribuir com ferramentas eficientes para a transferência de conhecimento de um indivíduo para outro e para explicitar conhecimento tácito. Depreende-se do acima exposto, que para a Eletronuclear , e as demais empresas da mesma indústria, a gestão do conhecimento não é só uma grande oportunidade de desenvolvimento estratégico, é uma necessidade vital. (p. 4)".

Diante disso, para evitar a perda de conhecimento e evitar a manifestação de suas conseqüências, tornou-se necessária uma ação corporativa consistente. Assim, a questão da preservação da tecnologia existente se tornou uma prioridade e a ELETRONUCLEAR tomou a iniciativa de instituir em 01/01/2001, o Projeto *"Determinação do Know-How Tecnológico da ELETRONUCLEAR"* (LEPECKI, 2002) com o objetivo de inventariar o "know-how" existente, identificar as lacunas de conhecimento nas atividades essenciais, e propor soluções para sanar tais lacunas.

10.2 O Projeto de Gestão do Conhecimento da Eletronuclear.

10.2.1 Considerações Iniciais

Com o trabalho de LEPECKI (2002) deflagrou-se uma série de iniciativas e ações organizadas que deram origem ao Projeto de Gestão do Conhecimento da Eletronuclear. Foi elaborada uma taxonomia do negócio da empresa, e feito um diagnóstico da situação atual e para os próximos cinco anos com todos os gerentes, quanto ao conhecimento (entenda-se "*know-how*") necessário para as atividades diretamente relacionadas com o negócio da empresa

Com o projeto *"Determinação do Know-How Tecnológico da ELETRONUCLEAR"* conduzido por LEPECKI (2002) foi elaborada uma taxonomia do conhecimento requerido pelo negócio da empresa, dando início à busca de uma forma de classificar o seu "*know-how*". Foi feito, à época, um diagnóstico da situação atual e para os próximos cinco anos com todos os executivos, quanto à questão do conhecimento empresarial. Existiam então várias sistemáticas de classificação de "*know-how*" na ELETRONUCLEAR, mas cada uma com uma finalidade específica. Em particular, elas estavam fortemente vinculadas à usina considerada (Angra 1 ou Angra 2). Assim, decidiu-se por criar uma sistemática nova, adaptada à realidade da

empresa. Para tanto, foram consultados Diretores, Superintendentes, Gerentes e especialistas, sendo o produto obtido fruto de uma atividade interativa ocorrida ao longo de sete meses (março a outubro de 2001). A Sistemática de Classificação de *Know-How* Tecnológico da ELETRONUCLEAR, assim obtida, está dividida em quatro grandes conjuntos:

- Engenharia de Projeto e Apoio à Operação,
- Realização Física do Empreendimento,
- Operação
- Atividades de Sustentação da Empresa.

Estes grandes temas foram sendo subdivididos em níveis sucessivos de detalhamento. O número de níveis foi deixado a critério de cada área, não havendo uniformidade quanto a este aspecto, e tendo-se chegado a um total de 557 itens, conforme relatado por LEPECKI (2002) e mostrado no ANEXO I – Resumo do Relatório Projeto Determinação do 'Know-How' Tecnológico da Eletronuclear.

Gerou-se então uma lista das atividades da empresa mais intensivas em conhecimento, já que ela foi produzida pela resposta à seguinte pergunta colocada para os gerentes: "*Que "know-how" sua área necessita para desenvolver suas atividades*?" As respostas
apresentadas tiveram o formato: "*É necessário know-how para* ...." Desta forma não se obteve diretamente um mapa do conhecimento, mas sim uma lista de atividades para as quais o conhecimento é um insumo essencial. Esta lista de 557 atividades transformou-se no ponto de partida para o mapeamento do conhecimento da empresa.

De acordo com OLIVEIRA (2) (2002), com base neste levantamento foram definidas, juntamente com a diretoria da empresa, as atividades críticas, aquelas que necessitam de ações imediatas. Foi deflagrado um mapeamento detalhado do conhecimento destas áreas, com o objetivo de se desenvolver um plano de ação. O passo seguinte será definir uma estrutura e um plano de ação para transformar a gestão do conhecimento de um projeto, em uma atividade contínua integrada aos processos de negócio da Eletronuclear, atendendo assim a resolução da 4ª Reunião Ordinária do

CNPE – Conselho Nacional de Política Energética, de 05/12/2001, e satisfazendo uma das condições para a retomada da construção da Usina Angra 3.

10.2.2 Metodologia Proposta

A ELETRONUCLEAR identificou oportunidades para várias iniciativas vinculadas ao processo de Gestão do Conhecimento, considerando o porte da empresa e muitas de suas atividades operacionais serem intensivas em conhecimento. Tal processo constituir-se-ia em um projeto mais amplo, sob o qual estariam vários outros projetos subordinados e integrados. Por exemplo, a reavaliação dos processos de geração de energia nuclear com relação à experiência e boas práticas internacionais, quanto a geração de rejeitos sólidos indica um projeto de *Melhores Práticas*. Por outro lado, desenvolver um programa de transferência de conhecimento, planejar a substituição de pessoal com maior antecedência indicavam a relevância de um projeto de *Páginas Amarelas*. De igual forma, fazia-se necessário avaliar os aspectos relacionados à tecnologia da informação para sustentar as atividades de conhecimento da empresa.

OLIVEIRA (2) (2002) propôs iniciar o projeto de Gestão do Conhecimento da Eletronuclear com a aplicação da metodologia desenvolvida pelo CRIE – Centro de Referência em Inteligência Empresarial – COPPE/UFRJ, baseando-se em CAVALCANTE & GOMES (2002) para desenvolvimento de um projeto de gestão do conhecimento, com as necessárias adaptações para o caso específico da empresa. Esta metodologia é composta das seguintes etapas (Figura 39):

- Etapa I: Análise da Empresa e do Ambiente de Negócios
- Etapa II: Identificação do Projeto
- Etapa III - Plano de Ação
- Etapa IV – Avaliação

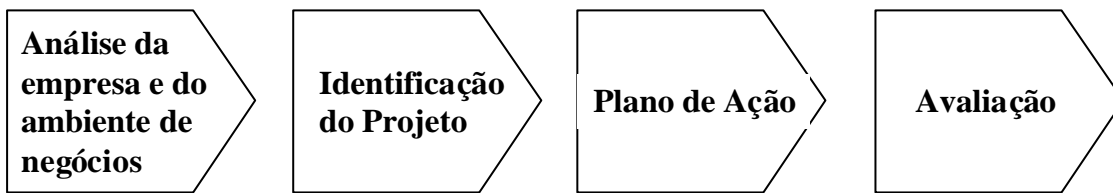



Figura 39: Etapas da Metodologia de Gestão do Conhecimento do CRIE
Fonte: CAVALCANTE & GOMES (2002)

10.2.3 O Projeto de Mapeamento do Conhecimento da Eletronuclear

Muitas oportunidades de projeto foram identificadas, mas era necessário priorizar. Essa metodologia foi, então, inicialmente aplicada em um projeto de Mapeamento do Conhecimento da Eletronuclear, do tipo construção de *Páginas Amarelas*. Esse tipo de projeto visa mapear o conhecimento interno da organização, descrevendo e avaliando os capitais humano, estrutural e de relacionamento, baseando-se na perspectiva de STEWART (1998) em que:

- Capital intelectual (ou humano): inclui conhecimento e competências individuais dos funcionários;
- Capital estrutural: inclui conhecimento ou competência coletiva, como processos, *know-how*, marcas e patentes, documentos;
- Capital de relacionamento: inclui o conhecimento e vantagens advindas dos relacionamentos entre funcionários com clientes, fornecedores, provedores de tecnologias, etc.

Foram identificados motivos para esse projeto, quais sejam:

- Dar continuidade natural ao que vinha sendo desenvolvido;
- Permitir criar a possibilidade de explicitar conhecimento tácito, através de ferramentas adequadas como, por exemplo, os Mapas Conceituais;
- Facilitar a captura de conhecimento explícito não documentado;
- Facilitar a criação de regras de busca para ferramentas de “Mineração de Dados” e “Mineração de Textos”, que seriam implementados em seqüência;

- Identificar a correlação entre especialistas, dentro e fora da empresa, com as atividades dos processos empresariais, facilitando a busca e o compartilhamento de conhecimento;
- Facilitar a aplicação de outros projetos de gestão do conhecimento, uma vez que os Mapas do Conhecimento já foram elaborados, e estarão sendo mantidos atualizados.

Para realizar o mapeamento do conhecimento com o projeto de páginas amarelas, foi estabelecida uma metodologia com o seguinte fluxo:

- Identificar os processos a serem mapeados e suas atividades;
- Identificar as pessoas chave e relacionar as mesmas às atividades;
- Identificar os capitais humano, estrutural e de relacionamento necessários para realizar as atividades;
- Identificar os capitais humano, estrutural e de relacionamento existentes para realizar as atividades;
- Desenvolver mecanismos de atualização e nivelamentos dos referidos capitais.

Os 5 passos dessa metodologia deve ser percorridos para cada uma das 557 atividades identificadas para a ELETRONUCLEAR no projeto *"Determinação do Know-How Tecnológico da ELETRONUCLEAR"* (LEPECKI, 2002), o que obviamente consome bastante tempo e está em curso na empresa. Para o levantamento do Capital Intelectual, procura-se responder a pergunta: *"Qual ou quais habilidades/competências são necessárias para realizar a atividade?"* Devem ser coletadas as seguintes informações por atividade:

- Capital intelectual – habilidades/competências necessárias
- Número de pessoas necessárias
- Nível de experiência requerido

De acordo com OLIVEIRA (2) (2002), para obter as informações acima, se faz necessário definir quais são os perfis profissionais necessários a execução da atividade, e quantas pessoas com cada perfil são necessárias. Para auxiliar nesta tarefa utiliza-se o modelo de classificação de competências que define habilidades básicas, específicas,

globais e universais, conforme mostrado na Figura 40, baseada na estrutura utilizada pela Microsoft e apresentada por DAVENPORT (1997) .

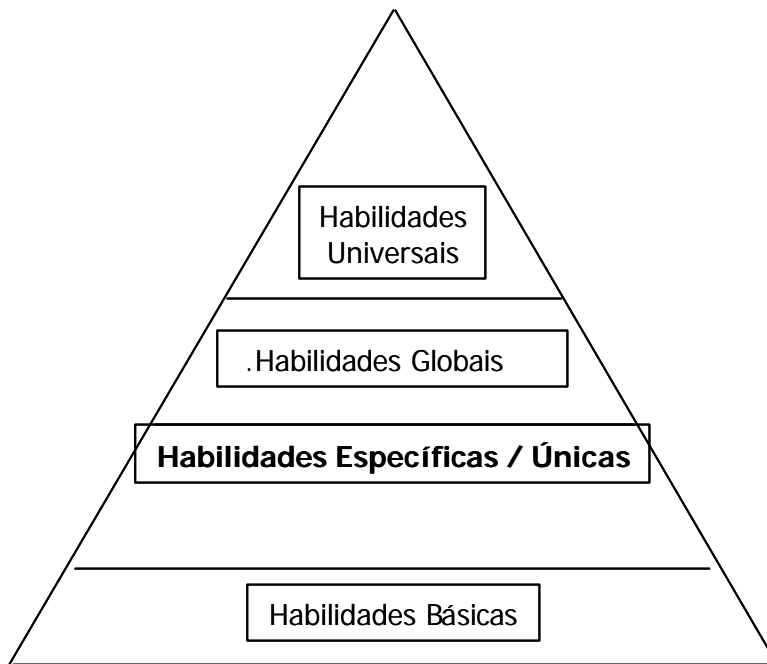



Figura 40: Estrutura genérica de competências de uma atividade
Fonte: DAVENPORT (1997)

Nesse modelo, conforme DAVENPORT (1997), temos as seguintes definições:

- As *Habilidades Básicas* são as chamadas competências básicas, ou seja: aquelas necessária para entrar neste tipo de atividade.
- As *Habilidades Específicas / Únicas* são competências avançadas aplicáveis a um tipo particular de tarefa.
- As *Habilidades Globais* são aquelas que todos os funcionários de uma determinada função, ou organização devem ter.
- As *Habilidades Universais.* são aquelas que todos os funcionários da empresa precisam ter.

Aplicando concretamente tais idéias em um exemplo para a atividade de um Engenheiro de Materiais termos, conforme OLIVEIRA (2) (2002) a Figura 41 abaixo:

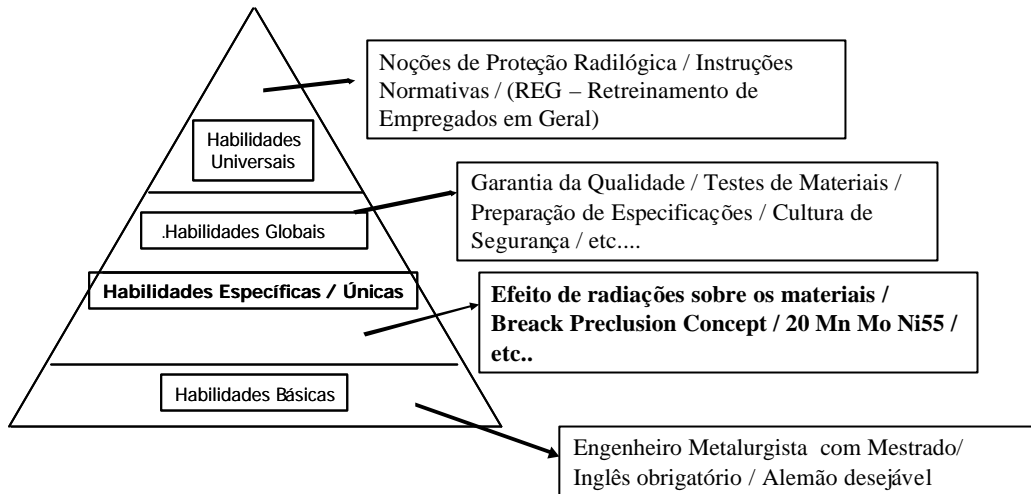



Figura 41: Estrutura de competências da atividade de Engenharia de Materiais
Fonte: OLIVEIRA (2) (2002)

Para o levantamento do Capital Estrutural, procura-se responder a pergunta: "*Qual ou quais estruturas/sistemas/métodos são necessários para realizar a atividade?*" Devem ser coletadas as seguintes informações por atividade:

- Especificação e intensidade requerida de cada estrutura, sistema, procedimento; documento, programa de computador, computador, rotina, etc;
- Nível de prioridade requerido.

Para o levantamento do Capital de Relacionamento, procura-se responder a pergunta: "*Qual ou quais relacionamentos são necessários para realizar a atividade?*"A metodologia do CRIE adotada como referência para este projeto estabelece que deve ser feito o mapeamento dos relacionamentos internos, e externos, estes últimos tanto com pessoas de fora da organização como com outras instituições. No caso da Eletronuclear entendemos ser conveniente no momento proceder apenas o levantamento dos relacionamentos externos, pelas seguintes razões:

- Existem muitos processos de alta complexidade gerando um alto número de interfaces, o que geraria uma rede de relacionamentos de difícil administração, pelo menos neste primeiro momento, e:

- As barreiras culturais a este levantamento, pelo menos no início do processo, podem prejudicar o andamento do projeto como um todo.

O Mapa de Relacionamentos poderá será gradativamente construído, com um processo de transformação cultural voltado para a gestão do conhecimento, já em curso. Para o Mapeamento do Capital de Relacionamento foram coletadas as seguintes informações sobre os relacionamentos com pessoas e entidades externas, importantes para a execução da atividade:

- Pessoa ou Instituição com quem há o relacionamento;
- Tipo de relacionamento com pessoas (pessoal ou profissional), e nível atual do relacionamento (excelente, bom, indiferente, ruim, péssimo);
- Tipo de relacionamento com instituições (formal – existe uma parceria com documentos formalizados ou informal – existe uma parceria de trabalho sem documentos formalizados), e nível atual do relacionamento (excelente, bom, indiferente, ruim, péssimo).

Por exemplo, podemos a apresentar a seguinte estrutura na Figura 42:

Atividade		
Elaboração de Conceito de Materiais e Especificações para os Componentes do Circuito primário		
Relacionamento Externo (Instituições)	**Tipo**	**Nível**
Westinghouse	2	4
Siemens	2	4
EPRI	2	4
TÜV	2	4

Figura 42: Exemplo de Mapeamento do Capital de Relacionamento Externo da Eletronuclear.

Fonte: OLIVEIRA (2) (2002)

Uma vez que se tenham os Mapas dos Capitais do Conhecimento devidamente armazenados em planilhas eletrônicas e bancos de dados, é possível uma série de consultas, conforme listado abaixo e de acordo com CAVALCANTE & GOMES (2002) na metodologia CRIE:

- Relatório do capital estrutural existente;
- Relatório do capital intelectual existente;

- Relatório do capital de relacionamento existente (através das pessoas-chave);
- Relatório do capital de relacionamento existente (através da organização);
- Buscas por pessoas-chave X atividade;
- Buscas por pessoas-chave X habilidade/relacionamento;
- Buscas por atividades X pessoa-chave;
- Relatório das lacunas de uma atividade;
- Relatório das pessoas-chave x lacunas/atividade;
- Relatório de capitais do conhecimento de uma pessoa-chave não aproveitados pela empresa;
- Relatório das lacunas do capital estrutural da organização;
- Relatório das lacunas do capital de relacionamento da organização.

10.2.4 Desafios Decorrentes do Mapeamento do Conhecimento da Eletronuclear

10.2.4.1 Fundamentos e Propostas

O processo de mapeamento do conhecimento deflagrou na Eletronuclear um conjunto subseqüente de ações no campo da gestão do conhecimento. Essas ações estão fortemente ancoradas nas perspectivas de STEWART (1998), que trata dos capitais intelectual, estrutural e de relacionamento, e mais especialmente na de NOVINS & ARMSTRONG (2002), descrita no item 4.3.2.2 da presente tese, que aborda a aplicação e transferência do conhecimento. Também merece menção a perspectiva de divisão entre conhecimento tácito e explícito, proposta por POLANYI (1966) e que permeia toda a discussão relativa aos papeis das áreas da empresa, especialmente as de recursos humanos (mais voltadas para o conhecimento intelectual / tácito) e de tecnologia da informação (mais voltada para o conhecimento estrutural / explicito).

OLIVEIRA (2) (2002), partindo da visão do conhecimento em termos de aplicabilidade e transferência de NOVINS & ARMSTRONG (2002), identificou e propôs para a ELETRONUCLEAR ações a serem tomadas para permitir, ou facilitar o processo de gestão do conhecimento. Essas ações são complementares à elaboração dos Mapas dos Capitais do Conhecimento. Tais ações são candidatas a entrarem no plano de ação para implementação da gestão do conhecimento na Eletronuclear. Com base no

trabalho *Choosing Your Spots for Knowledge Management*, dos referenciados autores, foi sugerido por OLIVEIRA (2) (2002):

a) Para gerenciar Conhecimento de Acesso Rápido:

Aplicação de ferramentas tecnológicas de *mineração de textos e mineração de dados* com ponto forte nas primeiras, já que nos repositórios de conhecimento explícito da ELETRONUCLEAR predominam os documentos textuais. A ELETRONUCLEAR tem a sua disposição grandes repositórios de conhecimento explícito, internos e externos a organização, que não são aproveitados em todo o seu potencial. O modelo proposto também deve apresentar mecanismos para melhorar o aproveitamento dos conhecimentos contidos nestes repositórios. Quanto mais poderosa for a ferramenta (de mineração de textos) na busca de conhecimento em repositórios não estruturados melhor será para a empresa. A chave para o sucesso neste caso é uma boa taxonomia. Os motores de busca disponíveis hoje são poderosos, mas só acessam documentos eletrônicos obviamente. Então também é fundamental a continuidade da implantação do processo de Gerenciamento Eletrônico de Documentos, praticamente concluído. Além disso, muitos dados importantes da operação e engenharia já se encontram em bancos de dados estruturados, já sendo de fácil captura.

b) Para gerenciar Conhecimento Geral

Uma ferramenta tecnológica importante na missão de gerenciar conhecimento é um *portal corporativo* com recursos de controle de documentos, para garantir que não serão acessadas informações obsoletas, principalmente com recurso de assinatura de documentos, ou seja: com disparo automático de e-mail para todos os assinantes da informação / conhecimento, sempre que surgir algo novo, ou houver exclusão de documentos importantes.

c) Para gerenciar Conhecimento Complexo

Devem ser desenvolvidos programas de treinamento com base nos "*gaps*" atuais e futuros identificados no mapeamento do conhecimento. Desenvolvimento de

planos de sucessão. A utilização do "*mentoring*" parece se mostrar uma forma bastante adequada por ser aceito na cultura empresarial. A maioria do pessoal mais experiente foi preparado desta forma, já que foi um método amplamente utilizado nos acordos de transferência de tecnologia. Aplicar técnicas avançadas de captura do conhecimento desenvolvido em uma atividade, como os mapas conceituais[85]. Esta ação pode ser desenvolvida em conjunto com o EPRI[86] e outras empresas de energia elétrica dos Estados Unidos que estão fortemente engajados no aperfeiçoamento do uso destes mapas[87]. O resultado obtido é explicitação de conhecimento tácito e troca mais eficiente deste tipo de conhecimento.

d) Para gerenciar o compartilhamento do conhecimento "um-para-um", e "um-para-muitos".

Este tipo de compartilhamento de conhecimento já está embutido nas ações tomadas nos itens anteriores.

e) Para gerenciar o compartilhamento do conhecimento "muitos-para-um".

As ações que necessitam ser realizadas para gerenciar o compartilhamento do conhecimento muitos-para-muitos também atenderão este caso, que se transforma em um caso particular daquele.

---

[85] Mapa conceitual é uma ferramenta para organizar e representar conhecimento. Um mapa conceitual deve representar as relações significativas entre conceitos, na forma de proposições. Proposições são dois ou mais conceitos, relacionados por palavras de modo a formar uma unidade semântica. Ele representa conceitos, organiza pensamentos, observa matizes de significados, resumir áreas de estudo, estimula o aprender como aprender e explicita a estruturas cognitivas. A sua teoria de fundo é construtivista. Representação gráfica em duas dimensões de um conjunto de conceitos construídos de tal forma que as relações entre eles sejam evidentes. Os conceitos aparecem dentro de caixas nos nós do grafo enquanto que as relações entre os conceitos são especificadas através de *frases de ligação* nos arcos que unem os conceitos. A dois ou mais conceitos, conectados por frases de ligação criando uma unidade semântica. As proposições são uma característica particular dos mapas conceituais se comparados a outros fatos similares como os mapas mentais. Os mapas atualmente tem sido implementados por softwares destinados a tal fim. O CMap, é um exemplo e é usado tanto por cientistas da NASA, que neles registram conhecimentos atuais sobre determinado planeta, quanto por crianças em escolas públicas dos EUA. Um bom detalhamento e apresentação de ferramenta é feito no site do *Institute for Human and Machine Cognition (IHMC)/University of West Florida* - http://cmap.ihmc.us. Acesso em 22.09.2006.

[86] *Energy Power Research Institute.* – Instituto de referência na área de energia elétrica que realiza pesquisas e estudos, elabora propostas e cria marcos de referência técnica para o setor e ao qual a ELETRONUCLEAR é filiada. É uma cooperativa de pesquisa e desenvolvimento na área nuclear.

[87] Visão apoiada no documento: *Capturing Undocumented Worker – Job-Knowledge – Overview an 2000 Status Report, September 2001. EPRI, Palo Alto, C:2001*

f) Para gerenciar o compartilhamento do conhecimento muitos-para-muitos.

Nesse ponto o foco principal sai da tecnologia da informação e passa para as pessoas. A forma de compartilhamento proposta não é usual ainda e as ferramentas que podem ser usadas para tal fim não possuem, por enquanto, grande histórico de sucesso, pelo menos no meio empresarial. Urge necessário atuar na cultura da empresa e preparar o pessoal para ver o compartilhamento de conhecimento por um outro prisma. Campanhas de comunicação devem ser empreendidas para que o processo flua com transparência desde o começo. Se a troca de conhecimento tácito pessoa a pessoa já encontra resistência, a troca coletiva tende a ser ainda mais complexa. Será, portanto, necessário planejar um processo de adaptação cultural da empresa conduzido por pessoal experiente e capaz. É preciso, além da transformação cultural, criar uma estrutura metodológica e tecnológica que propicie a troca de conhecimento complexo, na modalidade muitos para muitos. Uma ferramenta que vem sendo testada para esta finalidade é o mapa conceitual desenvolvido coletivamente, com apoio da tecnologia da informação. Na Eletronuclear há um projeto adequado para aplicação de mapas conceituais para compartilhamento de conhecimento que é o levantamento detalhado das bases de projeto das usinas, tarefa que só é possível através de uma troca coletiva de conhecimento sobre como as usinas foram projetadas e construídas.

A proposta de OLIVEIRA (2) (2002) redundou numa definição clara dos papéis a serem assumidos por diferentes instâncias na organização. Foram propostas quatro instâncias de atuação no processo de gestão do conhecimento, com papeis descritos a seguir.

a) Papel da Coordenação de Gestão do Conhecimento: (duas a três pessoas)

- Criar infra-estrutura tecnológica e cultural adequada
- Emitir Programa de Gestão do Conhecimento a cada ciclo de Planejamento Estratégico e coordenar sua implantação
- Zelar pelo alinhamento estratégico das ações de gestão do conhecimento

b) Papel da Administração de Recursos Humanos:

- Estabelecer juntamente com as áreas técnicas e operacionais os perfis necessários a execução das diversas atividades;
- Coordenar um processo de mudança cultural para criar um ambiente adequado a criação e o compartilhamento de conhecimento;
- Providenciar programa de treinamento para reduzir as lacunas identificadas no mapeamento dos capitais do conhecimento;
- Organizar plano de substituição do pessoal (plano de sucessão), e programas de treinamento no trabalho (*mentoring*);
- Desenvolver e implementar uma política de adequada de Recursos Humanos.

c) Papel das Relações Institucionais:

- Desenvolver um processo de comunicação e *"endomarketing"*, que mantenha o pessoal envolvido com os projetos de gestão do conhecimento;
- Desenvolver atividades que promovam a socialização, reduzindo as barreiras de comunicação entre as pessoas.
- Poderá vir a empreender atividades de monitoração social e política, em futuros projetos de inteligência empresarial.

d) Papel de Tecnologia da Informação:

- Preparação de planilhas eletrônicas e banco de dados para organização dos Mapas dos Capitais do Conhecimento
- Suporte na escolha e aquisição de softwares para mineração de textos, para elaboração de mapas conceituais e para gerenciamento eletrônico documentos;
- Prover a empresa de um portal corporativo;

10.2.4.2 Plano de Ação Para Mapeamento dos Capitais do Conhecimento – Projeto Piloto da Eletronuclear

O descrito no item 10.2.4.1 é um projeto amplo. Um plano de ação para o início do projeto, ou seja, o mapeamento dos capitais do conhecimento de um primeiro grupo

de atividades, em situação mais crítica (em seqüência ao descrito no ANEXO I) foi então proposto. O objetivo central foi o de realizar o mapeamento dos capitais do conhecimento de um primeiro conjunto de atividades, consideradas essenciais, ou seja de alta relevância estratégica e intensivas em conhecimento. O Plano de Ação propôs:

- Campanha de "endomarketing"
- Preparação das pessoas chave
- Elaboração dos perfis profissionais das atividades consideradas
- Preparação das planilhas e banco de dados a serem usados para registrar e trabalhar os dados coletados
- Mapeamento dos capitais do conhecimento necessários, e dos capitais estrutural e de relacionamento existentes
- Mapeamento dos capital intelectual existente
- Elaboração do plano de ação para os *gap's* juntamente com as áreas envolvidas
- Análise crítica do processo visando sua melhoria

Conforme OLIVEIRA (2) (2002), para avaliar o resultado obtido na implantação da gestão do conhecimento, a questão essencial de ser respondida é: "*qual o retorno esperado com os projetos*?" Já que é difícil quantificar os retornos provenientes do conhecimento, é preciso basear-se nos indicadores mais gerais de sucesso. Os indicadores de sucesso de projetos de Gestão do Conhecimento não são tão diferentes dos critérios que as empresas usam para medir o sucesso de outros tipos de projetos de mudança corporativa. Assim, segundo ou autor, poder-se-ia adotar os seguintes indicadores para avaliação de nosso processo:

- Resultados de melhoria de indicadores da WANO para o qual exista um projeto de gestão do conhecimento (Ex. - Volume de rejeitos sólidos – *benchmarking*)
- Número de atividades mapeadas por mês;
- Efetiva implementação do plano para eliminar "gaps" de conhecimento
- Número de consultas aos repositórios de conhecimento explícito
- Pesquisa de clima organizacional
- Efetivo cumprimento do Programa de Gestão do Conhecimento de cada ciclo de Planejamento Estratégico

10.2.4.3 Ações Empresariais Realizadas

Além do trabalho realizado e descrito em LEPECKI, 2002, já citado, a ELETRONUCLEAR deflagrou e vem conduzindo desde 2001 uma série de ações relativas ao processo de Gestão do Conhecimento, sendo que algumas merecem especial destaque e podem ser descritas e comentadas.

a) Execução de uma série de iniciativas relacionadas à Administração de Recursos Humanos, tais como:

- *Feedback* Técnico da Usina de Angra 2 para Angra 3.
- Desenho da Árvore de Competências para áreas técnicas.
- Projeto Determinação do *Know How* Tecnológico da ETN-2002 a 2004 para áreas técnicas e operacionais.
- Formação e retreinamento no Centro de Treinamento de Mambucaba (Angra dos Reis).

b) Execução de uma série de iniciativas relacionadas à Tecnologia da Informação, em consonância com o descrito no item anterior, quais sejam:

- Implantação da ferramenta *Sincronia*® de GED – Gerenciamento Eletrônico de Documentação Técnica (Programa de Gestão de Documentação);
- Implantação de ferramenta de portal corporativo, com aquisição da ferramenta *Verity*® para viabilizar o Portal de Experiência Operacional de Usinas;
- Implantação de Automação de Processos – *Workflow*;
- Busca e classificação e uso de conteúdo em documentos não estruturados (textos nos mais diversos formatos) ou *Unstructured Document Management (UDM)*. Num primeiro momento, uma resposta preliminar e significativa poderia ser gerada pela combinação das soluções de GED e de Portal, mas em um segundo momento, considerando a dimensão da base de dados e multiplicidade de usuários com diferentes interesses e necessidades, foi necessário deflagrar também os estudos necessários para implantação da tecnologia de mineração de textos, o que subsidia os estudos da presente tese.

c) Institucionalização do Processo de Gestão do Conhecimento na empresa, com a criação, em 2004, do Comitê de Gestão do Conhecimento. Isso foi realizado através dos normativos internos oficiais[88]. São objetivos desse comitê:

- Definir políticas e diretrizes para a implantação de projetos relativos à gestão do conhecimento - GC, no âmbito da empresa, como uma atividade permanente, com a participação de todas as u.o.s envolvidas, bem como, transferência das atividades implantadas, em caráter permanente, para as áreas gestoras;
- Fazer um planejamento bianual, com revisões trimestrais, das atividades de GC planejamento bianual de Gestão do Conhecimento;
- Promover campanhas internas de esclarecimento e de mobilização com o objetivo de equalizar conceitos e divulgar internamente a Gestão do Conhecimento.

Esse Comitê, composto por 6 integrantes de diferentes áreas da empresa e um consultor externo, realizou 27 reuniões entre dezembro de 2004 e agosto de 2005. Realizou a revisão de atividades já realizadas anteriormente à sua criação, fez contatos e reuniões com entidades externas, conduziu levantamento de ações normativas e organizacionais para suporte à Gestão do Conhecimento e desenhou proposta de programas e ações, dentre os quais se incluem:

- Programa de Sensibilização e Divulgação e criação da Semana da Gestão do Conhecimento;
- Projeto *Know How* para toda ELETRONUCLEAR;
- Árvore de Competência para toda ELETRONUCLEAR;
- Extensão das atividades do Centro de Treinamento para a Sede;
- *Avaliação Know How* em outras empresas;
- Estabelecimento de Ações Normativas e Organizacionais;
- Revisão dos Macro- Processos da ELETRONUCLEAR;
- Programa de Gestão do Conhecimento Implícito;

[88] RDE Nº 711.005/04, de 26 de outubro de 2004 e Circular Geral Nº 041/04, de 23 de novembro de 2004

- Desenvolvimento do Portal de Gestão do Conhecimento na ELETRONUCLEAR;
- Aprimoramento dos programas de Gestão de Documentação e Experiência Operacional – redundando em demanda por ferramenta de mineração de textos.

10.3 A Experiência Operacional na Eletronuclear

A expressão conhecimento organizacional é aplicada sobre todo o domínio de conhecimentos da organização, isto é, sobre todas as áreas e processos, incluindo também o conhecimento sobre o ambiente em que a organização está inserida. Muitas vezes, em processos de gestão do conhecimento, uma organização define prioridades e quais os domínios de conhecimento que pretende dar foco em suas ações. A experiência operacional é um subconjunto desse conhecimento organizacional. Como nome indica, tem foco definido: a dimensão operacional da organização – os aspectos vinculados à sua atividade fim. Na ELETRONUCLEAR, e em qualquer empresa responsável por gerar energia elétrica com fonte nuclear, a experiência operacional engloba os aspectos relacionados à operação, manutenção, comissionamento[89], paradas[90], segurança[91] e monitoramento de uma usina nucleoelétrica. Ela inclui fontes da própria organização e de fora dela, podendo então ter uma subdivisão entre externa e interna, que, porém, se complementam para viabilizar à adequada gestão do processo de operação e controle dos processos operacionais das usinas, com foco especial para

[89] Processo durante o qual componentes e sistemas da *usina* nucleoelétrica, tendo sido construídos e montados, são tornados operacionais, procedendo-se à verificação de sua conformidade com as características de projeto e com os critérios de desempenho e de aceitação (CNEN-NN-1.14. Resolução N°016 29/11/2001).

[90] Uma *Parada* é definida como um desligamento programado da *usina* para troca de *combustível nuclear*. Mas pode haver uma *paralisação forçada* que é a retirada de serviço da *usina* após a ocorrência de um *evento*, retirada essa necessária para adoção de ação corretiva conseqüente desse *evento*. Entende-se por *paralisação forçada* não apenas aquela que ocorre imediatamente após, mas também aquela que ocorre até o fim da semana seguinte à ocorrência do *evento*. (CNEN-NN-1.14. Resolução N°016 29/11/2001).

[91] *Segurança técnica nuclear* (ou simplesmente *segurança*) - conjunto de medidas de caráter técnico, incluídas no projeto, na construção, no comissionamento, na manutenção e na operação da *usina*, visando evitar a ocorrência de *incidentes* e *acidentes* ou minimizar suas conseqüências. Um *Incidente é um evento* sem dano significante à *usina* e/ou aos trabalhadores e ao público, mas significante em relação à *segurança* da *usina*. Pode incluir desligamentos não planejados, *paralisações forçadas* e violação de *condições limites para operação.* Um *Acidente* - desvio inesperado e significativo das condições normais de operação de uma instalação, incluindo ocorrências previstas, *acidentes postulados* ou *acidentes severos*, tais que possam resultar em danos à propriedade ou ao meio ambiente ou ainda em exposições de trabalhadores ou de indivíduos do público acima dos limites primários de dose equivalente estabelecidos. (CNEN-NN-1.14. Resolução N°016 29/11/2001).

apoiar as atividades de monitoramento e controle de eventos[92]. Informações e conhecimentos providos por fornecedores de equipamentos insumos e tecnologias para uma empresa operadora, bem como as informações de avaliações e testes diversos também são considerados. Uma representação desse processo, conforme indicado pela IAEA – *International Atomic Energy Agency* é apresentada na Figura 43. Portanto, temos uma Experiência Operacional Interna – EOI e uma Experiência Operacional Externa – EOE. O desejável para uma empresa como a ELETRONUCLEAR é procurar manter as usinas em funcionando em *operação normal* - (ou *condição normal de operação*) – operação que inclui todas as condições e *eventos* que são previstos ocorrer no curso da operação pretendida, quando realizada sob controles administrativos e de acordo com procedimentos especificados, dentro das *condições limites para operação.*

Uma forte motivação para o adequado gerenciamento do conhecimento, tanto no nível da EOI quanto da EOE é a minimização de riscos na operação, uma vez que um indesejado acidente nuclear, caso ocorra, poderia colocar em risco a credibilidade de todas as indústrias espalhadas no mundo todo. Por isso, as organizações operadoras de usinas ao redor do mundo, agências reguladoras, centros de pesquisas e de capacitação buscam promover o compartilhamento de conhecimentos sobre todos os aspectos da operação. Esse compartilhamento, concretamente significa não só um relacionamento entre profissionais das empresas que atuam no setor, mas, especialmente, o compartilhamento de informações, sobretudo sob a forma de documentos textuais. A base de conhecimento é resultado, portanto da contribuição de diversas empresas. O foco principal da presente tese é sobre a experiência operacional da ELETRONUCLEAR, em consonância com as prioridades que a empresa assumiu, desde 2001, pelos motivos e conjuntura já explanados no item 10.1 do presente trabalho.

[92] *Evento operacional* (ou simplesmente *evento*) é uma *falha* ou uma seqüência de *falhas* relacionadas entre si. (CNEN-NN-1.14. Resolução N°016 29/11/2001).

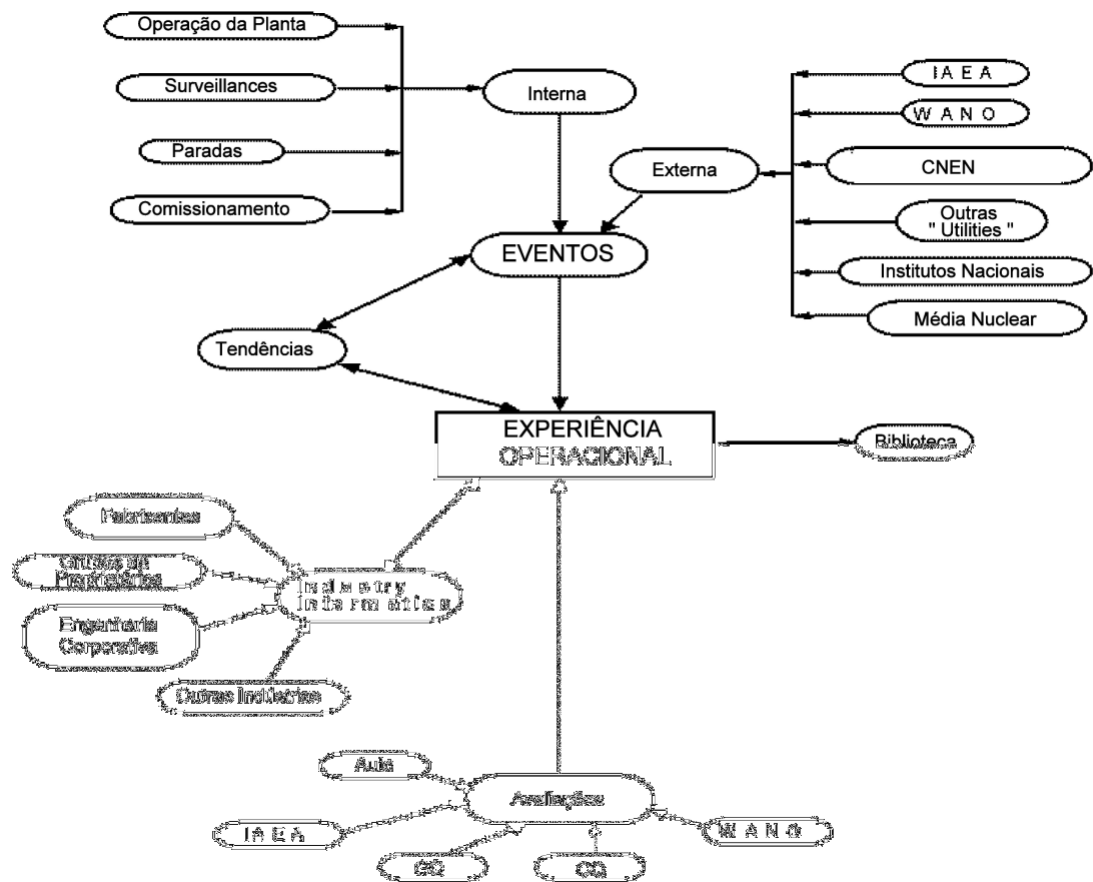



Figura 43: Processo de experiência operacional como recomendado pela IAEA

Fonte: OLIVEIRA (2) 2006

A experiência operacional, como um segmento do conhecimento organizacional da ELETRONUCLEAR tem, baseando-se na análise de POLANYI (1966) uma dimensão tácita e uma dimensão explicita, tanto para a EOI quanto para a EOE. A dimensão tácita é caracterizada pela experiência acumulada na prática de anos de operação das usinas de Angra I e II e pelos contatos e trocas de experiências, pesquisas e estudos que a empresa conduziu desde a década de 70 com profissionais, empresas operadoras, organismos reguladores e centros de pesquisa em todo o mundo. A dimensão explicita da experiência operacional está embutida dados diversos registrados em sistemas transacionais da empresa, em normas, procedimentos, relatórios diversos, documentos de projeto, engenharia, operação, manutenção e segurança das usinas todos gerados internamente. Inclui também, no caso da EOE, grandes bases de documentos (em sua maioria textuais) *importados/recebidos* de bases de outras empresas operadoras, organismos reguladores e centros de pesquisa no Brasil e em todo o mundo, tais como:

- AIEA – Agência Internacional de Energia Atômica
- CNEN - Comissão Nacional de Energia Nuclear
- WNA – Word Nuclear Association (Associação Nuclear Mundial)
- WANO (*Word Association of Nuclear Operators*)
- INPO (*Institute of Nuclear Power Operators*).
- NRC – *United States Nuclear Regulatory Commission*
- VGB – Instituição européia, sediada em Essen, na Alemanha e que atua na área nuclear.

Exemplificando e detalhando um pouco os domínios da EOI e da EOE, com base em OLIVEIRA, FILHO & SOUZA (2004) podemos dizer que eles dizem respeito a:

- a) Operação de Plantas Termonucleares (ANGRA 1 ANGRA 2) envolvendo EOI e EOE em:

  - Condução do fluxo de Operação para Geração de Energia;
  - Manutenção;
  - Engenharia de Sistemas;
  - Engenharia de Apoio à Operação
  - Segurança Ambiental
  - Segurança Nuclear
  - Treinamento
  - Experiência Operacional Interna e Externa
  - Materiais
  - Planejamento (paradas, testes, etc)
  - Comercialização
  - Qualidade e Licenciamento

- b) Engenharia, Construção e Suprimentos (*Engineering, Procurement & Construction – EPC*) implicando atualmente nos desafios de modificações de projetos das plantas em operação e modernizaçõe, troca do gerador de vapor da usina de Angra 1, construção do depósito definitivo de rejeitos e construção da Usina de angra 3. Tais conhecimentos envolvem, EOI e EOE:

  - Projeto Básico
  - Projeto de Detalhamento
  - Gerenciamento
  - Licenciamento
  - Planejmanto e Controle
  - Suprimentos
  - Contrução
  - Montagem
  - Comissionamento
  - Garantia e Controle da Qualidade.

Para concluir, devemos lembrar que relatórios diversos de operação, gerados no fluxo das atividades e do funcionamento das usinas são também considerados como experiência operacional, no caso, interna. Muitos relatórios são de confecção obrigatória, conforme determinação de agentes reguladores do setor, havendo normas para geração dos mesmos, tais como na Norma CNEN (2001)[93], apresentada no ANEXO II do presente trabalho. Há dois grandes grupos desses relatórios, quais sejam: as notificações de eventos e os relatórios rotineiros, todos regulados pela citada norma.

As notificações de eventos ocorrem com base na relevância do mesmo, classes de significância e suas conseqüências, grau de segurança, prazos de comunicação, dentre outros aspectos. Tais relatórios são gerados, sob a forma de documentos textuais eletrônicos e comunicados às instâncias competentes, internas ou externas à empresa,

[93] Relatórios de Operação de Usinas Nucleares. Norma tem por objetivo estabelecer os requisitos do programa de notificações de *eventos* significativos e dos relatórios de operação de *usinas nucleoelétricas* exigidos pela *Comissão Nacional de Energia Nuclear* (*CNEN*). Os relatórios decorrentes de disposições desta Norma devem ser endereçados à *CNEN*.

conforme o caso. A Eletronuclear tem hoje uma base significativa dos seguintes relatórios, cada um com sua respectiva estrutura:

- Relatórios de Evento – RE
- Relatórios de Desvio Operacional – RDO
- Relatórios de Evento Significante – RES
- Relatórios ORO – Ocorrência de Relato Obrigatório

Há também os chamados Relatórios Rotineiros, especificados em CNEN (2001), quais sejam:

- *Relatório Anual de Operação (RAO)* - relatório rotineiro, abrangendo a operação da *usina* durante o ano civil recém-findo.
- *Relatório Final de Análise de Segurança (RFAS)* - relatório que deve ser submetido à *CNEN* a cada pedido de *Autorização para Operação Inicial,* conforme previsto na norma CNEN-NE-1.04.
- *Relatório de Parada (RP)* – relatório emitido após cada *parada* da usina, abrangendo as atividades previstas e aquelas desenvolvidas durante esse período, assim como sua preparação e a experiência adquirida.
- *Relatório Mensal de Operação (RMO)* - relatório rotineiro de operação, abrangendo a operação da *usina* durante o mês civil recém-findo.
- *Relatório de Operações Iniciais (ROI)* - relatório resumido sobre o *programa de testes de partida*, incluindo testes de criticalidade, operação a baixa potência e elevação de potência até atingir o valor nominal.
- *Relatório do Programa de Monitoração Ambiental Radiológico Operacional* – apresenta os resultados dos parâmetros indicados no *PMARO*, no período a que se relaciona, por meio de amostras ambientais e com base nos caminhos críticos de transferência de radionuclídeos para o ser humano.
- *Relatório do Projeto Nuclear e Termohidráulico (RPNT)* - relatório que contém o projeto nuclear e termohidráulico, com sua respectiva *análise de segurança*, correspondente a cada configuração do núcleo do *reator*.

Diante do exposto, podemos concluir que a Experiência Operacional Interna e Externa, constituem, de certa forma, o cerne do conhecimento relativo à atividade fim

da ELETRONUCLEAR, justificando, portanto, esforços e investimentos no sentido de valoriza-lo, organiza-lo e dinamizar sua aplicação no contexto da empresa.

10.4 Sistemas e Tecnologias da Informação em Gestão do Conhecimento na Eletronuclear

10.4.1 Visão Geral

Um projeto de Gestão do Conhecimento, a partir do exposto na revisão de literatura, pode assumir diferentes matizes, focos, objetivos e prioridades. Pode também estar vinculado a diferentes estratégias e públicos organizacionais. Em um projeto como esse, coloca-se o imperativo da interação do capital intelectual da empresa com as soluções de tecnologia da informação adotadas. Há realmente esforços institucionais, humanos e tecnológicos envolvidos em um programa de Gestão do Conhecimento, como o da ELETRONUCLEAR. A presente tese procura enfocar as questões de ordem tecnológica, no campo da mineração de textos aplicada às bases documentais da empresa. Antes de entrar especificamente no foco principal, é relevante mapear o contexto tecnológico da empresa, sobretudo nos campos adjacentes aos de estudo do presente trabalho.

A ELETRONUCLEAR tem sede no Rio de Janeiro, sendo proprietária e operadora da Central Nuclear Almirante Álvaro Alberto, em Angra dos Reis, onde estão situadas as usinas de Angra I e II. Possui uma infraestrutura ampla, para suportar os desafios empresariais a que se propõe, com sistemas transacionais para administração da empresa e controle da operação de geração de energia. Opera com um modelo de computação distribuída, com redes internas de comunicação. Adota tecnologias contemporâneas para seu modelo de gestão tal como a aplicação ERP baseado no aplicativo R3 da SAP. Os dados transacionais são, como em qualquer outra empresa, muito importantes para a gestão financeira, contábil e de pessoal, dentre outras áreas. Entretanto, pela natureza da atividade, as bases documentais exercem particular importância na empresa e passaram a merecer tratamento cada vez mais avançado na organização em que a Gestão Eletrônica de Documentos o uso de Portais Corporativos e, mais recentemente, a tecnologia de Mineração de Textos passaram gozar de especial atenção da ELETRONUCLEAR.

Na perspectiva de OLIVEIRA, FILHO & SOUZA (2004) a interação entre Gestão do Conhecimento e Tecnologia da Informação na Eletronuclear pode ser representada pela figura 44 abaixo apresentada, em que dados geram informação, que por sua vez sustentam a gestão do conhecimento, alimentando as competências.

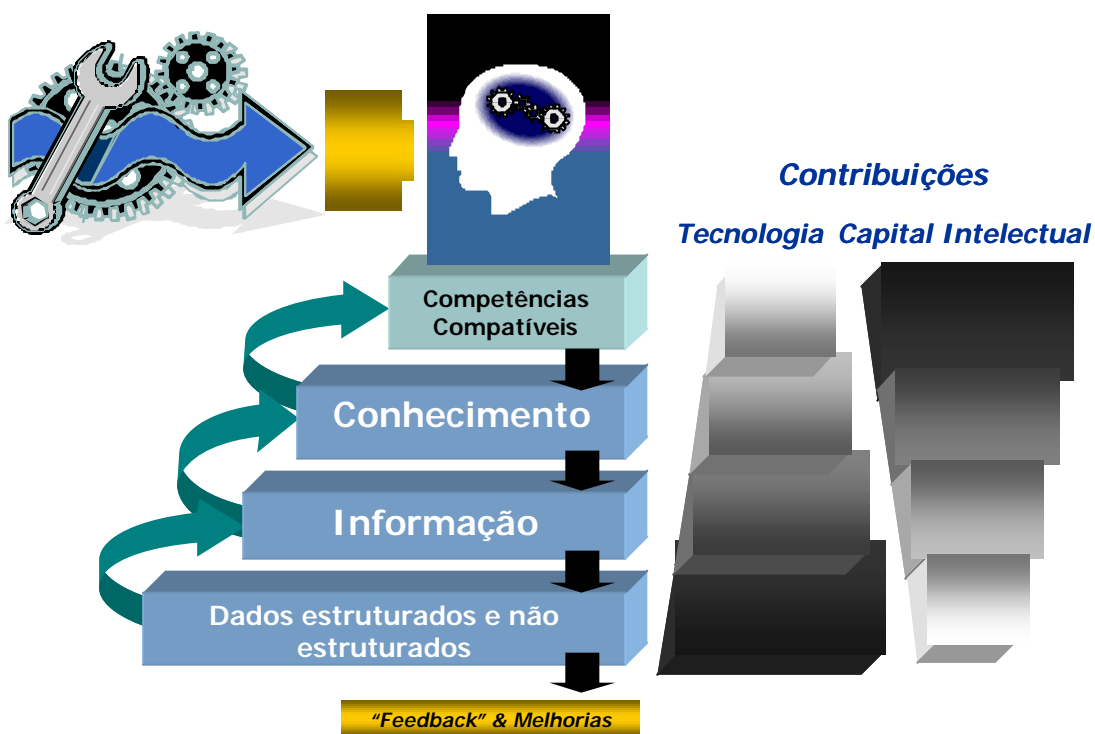



Figura 44: Interação do Capital Intelectual com Soluções de TI
Fonte: OLIVEIRA, FILHO & SOUZA (2004)

10.4.2 Gestão Eletrônica de Documentos, Automação de Processos e Portal Corporativo em Gestão do Conhecimento na Eletronuclear

Sendo o foco principal de atenção para a gestão do conhecimento os documentos textuais, sob as mais diversos formatos, como objetos portadores de informação potencial para diferentes grupos de usuários, foi necessário adotar visões do paradigma da recuperação da informação, descrito no capitulo 5 da presente tese. Em LEPECKI (2002) e no item 10.2.4.3 acima descritos foram relatadas as iniciativas, as ações empresariais realizadas para a Gestão do Conhecimento na empresa. No campo da Tecnologia da Informação e no âmbito de um projeto amplo de gestão do conhecimento, portanto, foi deflagrado na empresa, a partir de 2003, um processo englobando uma

série de ações integradas, focadas no tratamento e manejo da informação para apoio à gestão do k*now-how* tecnológico da Eletronuclear, quais sejam:

- Gerenciamento Eletrônico de Documentação Técnica - GEDT
- Automação de Processos - *Workflow*
- Busca e classificação de conteúdo não estruturado - *Unstructured Document Management* (*UDM*)

Cumpre salientar que esse processo visa, sobretudo contemplar a imperativa necessidade de gerencia da Experiência Operacional da ELETRONUCLEAR. A Gerência de Engenharia de Apoio – GEA.T assumiu a responsabilidade pelo processo[94]. De acordo com ela, os desafios atuais relacionados a gestão do conhecimento englobam:

- Melhor utilização dos repositórios de conhecimento disponíveis – Experiência Operacional
- Disponibilizar os documentos de engenharia e operação em meio eletrônico sem perda do "Controle da Configuração da Planta"
- Captura do conhecimento tácito de especialistas
- Levantamento de Árvore de Competências

O ambiente pré-projeto revelou problemas a serem superados. Em primeiro lugar, os sistemas estão preparados para processar dados, mas não para gerir conhecimento. Em segundo lugar, a Informação encontra-se dispersa por inúmeros computadores e sistemas. Em terceiro lugar, multiplicam-se as fontes de informação com conteúdos e formatos diversificados. Por fim, constata-se que uma grande quantidade de conhecimento disponível não é utilizado de forma eficaz devido a dificuldade de acesso. A Eletronuclear tem, no seu conjunto, uma significativa base de

[94] Equipe composta por *Luiz Celso Oliveira*, *Coordenador de Experiência Operacional Externa - EOE*, *Darcy dos Santos Filho*, *Coordenação da Implantação do GEDT*; *Marcos Roberto Vieira de Souza*, *Suporte em TI.* Essa equipe apresentou esse trabalho no V Seminário Nacional da Gestão da Informação e do Conhecimento no Setor de Energia Elétrica - SINCONEE, realizado em 30/06/2004 em Belo Horizonte MG. Tal trabalho é fonte das informações internas da empresa sobre o presente programa, descrito no presente item.

dados relacionais, bem como um grande acervo de documentos textuais, envolvendo documentação técnica, relatórios, projetos, dentre outros.

Considerando o contexto acima descrito, de acordo com OLIVEIRA, FILHO & SOUZA (2004), a opção da Eletronuclear foi a de implantar uma solução de GEDT (Gerenciamento Eletrônico de Documentos Técnicos) com dispositivo de *Workflow* (automação de processos), integrada com uma ferramenta de busca e classificação de conteúdo estruturado e não-estruturado, encapsulada em ambiente de Portal. Assim sendo, os usuários "consumidores de conteúdo" poderiam ter o atendimento dinamizado às suas necessidades de informação pela proposta da solução integrada *GEDT-Workflow-Portal*. Esse trabalho encontra-se em fase avançada de desenvolvimento na Empresa, podendo ser representado pela figura 45.

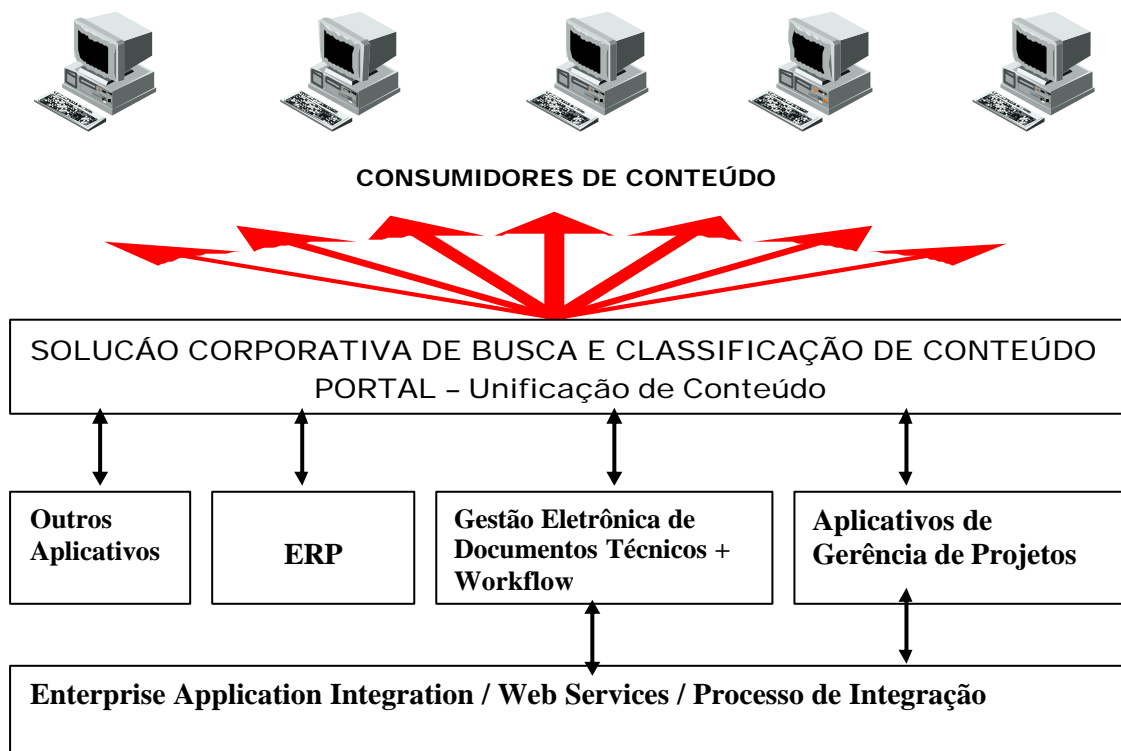



Figura 45: Gestão Eletrônica de Documentos, Automação de Processos e Portal Corporativo em Gestão do Conhecimento na Eletronuclear
Fonte: OLIVEIRA, FILHO & SOUZA (2004)

Inicialmente foram adotadas diretrizes e métodos para captura de conhecimento de grande valor ainda não-documentado. Isso envolveu os seguintes passos:

- Identificação de "*experts*" e de conhecimento de valor
- Compartilhamento, armazenagem, recuperação e apresentação do conhecimento concernente.
- Identificar e levar em consideração a infra-estrutura existente ou planejada.
- Implantação do plano.

A lógica da definição da solução adotada é a implantação de estratégias tecnológicas de recuperação da informação, com busca e classificação de conteúdo estruturado e não-estruturado e que oferece tecnologia diferenciada em escalabilidade[95] e velocidade de consulta. Realmente, a especialização técnica da área nuclear gera inevitavelmente um legado de sistemas especialistas com repositórios isolados. Este tipo de ferramenta capacita usuários "descobrirem conhecimento", antes isolado em tais repositórios, oferecendo um mecanismo de "*knowledge delivery*", evitando re-trabalho e reduzindo erros. Diante disso, entende-se que uma solução GEDT oferece funcionalidades especialistas no controle de documentação de engenharia em todo seu ciclo de vida, prevendo integração com ferramentas de gerenciamento de projetos, aplicativos CAD, gestão de referências, entre outras, que são essenciais para garantir a configuração da planta. A integração do GEDT com a ferramenta de busca e classificação de conteúdo (Portal) possibilita a consulta e visualização da ultima revisão (aprovada) de um documento técnico, sem necessitar que o usuário seja capacitado na utilização do GEDT para exercer esta função. Os custos de licenciamento para simples consultas a base documental são reduzidos significativamente, por não exigirem obrigatoriamente licenças do GEDT, que normalmente apresentam um custo maior de licenciamento do que adotar uma solução corporativa de busca e classificação de conteúdo.

O intuito principal da presente proposta é viabilizar a difusão do conhecimento em usinas nucleares a partir da implementação do conceito do EPRI de "PIM - *Plant Information Management*", representado na figura 46. Por essa perspectiva, a tecnologia de portal corporativo é o agente de difusão do conhecimento, na medida em que articula diferentes blocos de conteúdos de informações entre múltiplos usuários que são

[95] Em Tecnologia da Informação, a expressão indica a possibilidade de crescimento de uma aplicação de acordo com a necessidade de um ou de uma comunidade de usuários.

“consumidores de conhecimento”. Assim, as estruturas de dados podem ser consolidadas e partilhadas.

Figura 46: Conceito do *EPRI : “PIM - Plant Information Management*
Fonte: OLIVEIRA, FILHO & SOUZA (2004)

10.4.2.1 Portal de Experiência Operacional e Informação Tecnológica

Um projeto de Gestão do Conhecimento, a partir do exposto na revisão de literatura, pode assumir diferentes matizes, focos, objetivos e prioridades. Pode também estar vinculado a diferentes estratégias e públicos organizacionais. Em um projeto como esse, coloca-se o imperativo da interação do capital intelectual da empresa com as soluções de tecnologia da informação adotadas. Na seqüência do item anterior, é importante apontar o uso de tecnologia da informação como suporte ao processo de Gestão do Conhecimento, com a implantação da ferramenta de busca e classificação de conteúdo, ou seja: um portal corporativo que permita o compartilhamento de conhecimento. A análise do uso de portais deve estar alinhada às estratégias organizacionais e também ao tipo de portal implantado. Cada vez mais está difícil

dissociar Portal Corporativo da Gestão de Conhecimento, já que o portal não é somente uma implementação tecnológica. Assim, sua adoção deve considerar os direcionamentos tecnológicos e de negócios que são de interesse para a organização. De acordo com OLIVEIRA (2) (2006), o Portal realiza um processo de coleta e indexação e executa uma classificação a partir dos índices, criando taxonomias e categorias, com uso de árvores de conhecimento e desenvolve um processo para busca e navegação a partir desses índices criados, com uso de filtros e metadados. O processo de busca e navegação básicas da solução de portal é apresentado na figura 47 abaixo.

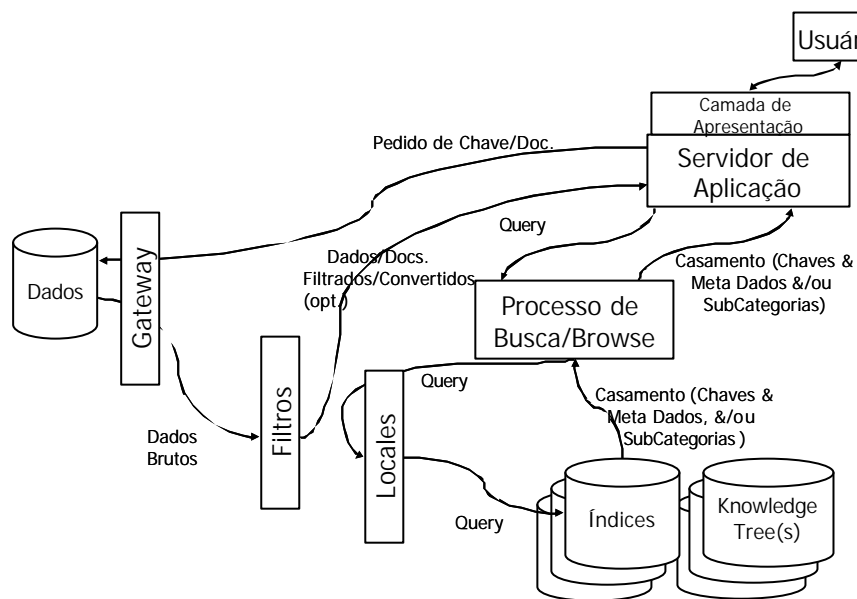



Figura 47: Processo de Busca e Navegação Básicas da Solução de Portal
Fonte: OLIVEIRA (2) (2006)

Possui recursos para pesquisas estruturadas e em texto completo, refinamento da seleção através de pesquisa paramétrica, operadores avançados e baseados em regras, navegação em documentos grandes e complexos com suporte para mais de 200 tipos de documentos diferentes, permitindo a visualização dinâmica dos documentos. Admite possibilidade de personalização no uso, com interface *web "customizável"*.

A ferramenta de portal permite lidar com diferentes comunidades de usuários com personalização baseada em papéis, direitos de acesso e preferências. O "*ranking*" permite que os usuários encontrem primeiro as informações que mais desejam. Há uso de agentes pessoais para receber notificação das modificações ocorridas no conteúdo do portal. O alcance é corporativo com a possibilidade de milhares de usuários simutâneos

que acessam centenas de milhares de documentos. Portanto, de acordo com a especificação previamente definida, essa ferramenta deve:

- Possibilitar buscas simultâneas em repositórios com conteúdos estruturados e não-estruturados;
- Oferecer mecanismo *'drill-down'* para navegar entre as categorizações de conteúdo.
- Apresentar resultados de busca relevantes para o perfil do usuário;
- Viabilizar o ambiente colaborativo através de redes sociais possibilitando recomendação de conteúdo entre usuários;
- Busca paramétrica através das categorizações de conteúdo, possibilitando selecionar filtros compostos de valores numéricos e períodos determinados por intervalos de datas;
- Busca federada: consolidar resultados de busca e categorização em portais externos que tem seus próprios mecanismos de busca;
- Possibilitar a definição de regras de negócio para automaticamente categorizar novos conteúdos, sem necessitar redefinir as taxonomias configuradas;
- Respeitar as permissões de acesso, definidas nos repositórios estruturados e não-estruturados.

Tal solução a ser adotada deve atuar em três camadas: apresentação, infraestrutura (ou fundação do portal) e agregação de conteúdo. Na Eletronuclear, a ferramenta escolhida para implementação foi o "*Verity*®", atualmente pertencente à empresa "*Autonomy*®" (EUA)[96]. Essa ferramenta encontra-se em fase final de implantação na empresa e as bases que agrega já se encontram disponíveis para os usuários internos da empresa. O ANEXO III transcreve uma página web da empresa *Autonomy*® com relato do caso da ELETRONUCLEAR.

[96] A empresa *Verity*, fornecedora de tecnologia de portais para a Eletronuclear, foi adquirida pela *Autonomy em dezembro de 2005.* Maiores informações em www.autonomy.com

10.4.2.2 Implantação da Gestão Eletrônica de Documentação Técnica e *Workflow*

A utilização de tecnologia de Gestão Eletrônica de Documentação Técnica – GEDT, integrada ao *Workflow* tem como objetivos:

- Gerenciamento de 700.000 documentos de engenharia.
- Gerenciamento e controle automático dos processos que envolvem documentação técnica;
- Velocidade e precisão na localização de documentos;
- Aumento de produtividade;
- Otimização dos procedimentos para tomada de decisões
- Possibilidade de "customização" de indexação para acesso a documentos;
- Redução de custos com cópias;
- Otimização na utilização de outras ferramentas, tais como: *ERP*, Aplicativos de Gerenciamento de Projetos, Aplicativos de gerenciamento de manutenção e parada de plantas industriais, dentre outras soluções de âmbito corporativo.

Na Eletronuclear a ferramenta escolhida para implementação da Gestão Eletrônica de Documentação Técnica, agregada ao *Workflow* foi o software brasileiro da empresa Sincronia®, que leva o mesmo nome. O "*Sincronia®*" publica documentos para indexação por ferramentas de busca e categorização de conteúdo, possibilitando consultas da base documental em toda organização. Na Eletronuclear esse software realiza o gerenciamento da documentação técnica das Usinas de Angra 1 e Angra 2, faz a automação dos processos utilizados no controle da configuração da planta e viabiliza a integração com solução de busca e categorização de conteúdo (*Enterprise Search and Content Categorization* - ESCC) em ambiente de portal.

As funcionalidades da ferramenta a ser adotada, englobam:

- Oferecer facilidades para integração com ferramenta de busca e classificação de conteúdo.
- Padronização de conteúdo para formato Adobe PDF.
- Acervo documental protegido contra acessos não-autorizados.

- Notificação antecipada de atrasos na emissão dos documentos.
- Integração com ferramentas de planejamento.
- Definição dos estados de um documento e permissões de acesso por nível funcional
- Modelagem gráfica e otimização contínua dos fluxos documentais.
- Acesso via *browser* (*thin-client*) em ambiente Web
- Gerenciamento das dependências de conteúdo entre documentos (*plant configuration management*)

10.4.2.3 Resultados Almejados com Portais, GEDT e *Workflow*

Dentre os benefícios esperados com a adoção de tais soluções tecnológicas, podemos citar alguns, sendo os principais vinculados ao processo de integração de processos informações e melhoria da disponibilidade da informação ao usuário. Essa consideração deve ser feita no âmbito da Gestão da Experiência Operacional da empresa. Do ponto de vista estratégico, temos: melhorar a identificação de oportunidades, adequar melhor as respostas dos sistemas de informações, controlar situações e otimizar recursos expressam benefícios gerados. Isso vem acrescido de outros ganhos, tais como: possuir conteúdos articulados, possuir eventos referenciados e informações consolidadas, além de garantir bons processos de interação entre conhecimentos, informações, pessoas e unidades organizacionais. São aspectos que melhoram as respostas da organização às suas necessidades.

11 MINERAÇÃO DE TEXTOS NA ELETRONUCLEAR

11.1 Considerações Preliminares

Inicialmente, embora a Eletronuclear tenha considerado a possibilidade de uso da mineração de textos como tecnologia de ponta para empresas de geração de energia e a partir de propostas de seu Projeto de Gestão do Conhecimento, conforme descrito no Capítulo 10, tal proposta havia sido postergada tendo em vista seus elevados custos para aquisição de ferramenta específica e não detenção de *know-how* tecnológico nessa área por parte da empresa. O presente trabalho de doutorado mudou esse quadro ao acenar com a possibilidade de introdução de *software* brasileiro de mineração de textos na empresa. Os primeiros contatos do autor da presente Tese com a Eletronuclear aconteceram a partir do **V SINCONEE –** Simpósio de Gestão do Conhecimento do Setor de Energia Elétrica, realizado em Belo Horizonte, em meados de 2004, a partir da interlocução com o Sr. Luiz Celso Oliveira, já citado nesse trabalho. A partir daí, e do estudo do documento elaborado por OLIVEIRA, FILHO & SOUZA (2004) várias visitas aconteceram à Central Nuclear Almirante Álvaro Alberto, em Angra dos Reis – RJ, para realização da Tese em questão.

A proposta de uso da tecnologia de mineração de textos na ELETRONUCLEAR já tinha sido identificada por OLIVEIRA (2) (2002) e sugerida por organismos de pesquisa da área de energia elétrica, como o EPRI para aplicação sobre a Experiência Operacional. Para esse contexto de aplicação, visa ampliar ainda mais os resultados e possibilidades do projeto Portal-GEDT-*Workflow* já deflagrado e descrito no capítulo anterior, criando novas possibilidades para os usuários na utilização das bases de documentos textuais relativas à Experiência Operacional da Empresa e existentes na empresa. Tais bases incluem documentos gerados por ela própria (Experiência Operacional Interna) ou documentos "importados" de outras bases externas, especialmente no caso da Experiência Operacional Externa. Podemos considerar que tais bases configuram o conhecimento explícito da ELETRONUCLEAR sobre atividades operacionais da empresa e sobre o qual o uso de tecnologias de informação pode apresentar bons resultados. As atividades administrativas e de apoio (como marketing, finanças e recursos humanos) também poderiam ser alvo de processo de descoberta de conhecimentos em suas bases textuais.

11.2 Caracterização da Necessidade de Mineração de Textos

O projeto Portal-GEDT-*Workflow* possui grande relevância para a ELETRONUCLEAR e recebeu investimentos importantes por parte da empresa. Entretanto, como toda tecnologia, possui seus limites e algumas importantes necessidades de informação não puderam ser contempladas por ele, mesmo atuando dentro do paradigma da recuperação da informação. Pretende-se com o uso de mineração de textos a realização do tratamento inteligente dos textos digitalizados, melhorando significativamente consultas, pesquisas, estudos técnicos e qualquer trabalho que venha a exigir um tratamento e análise avançada da informação, como suporte à gestão do conhecimento.

A ELETRONUCLEAR é uma empresa de grande porte e possui internamente diferentes comunidades de usuários de informação, em áreas administrativas e operacionais, cada uma com interesses e necessidades específicas. Opera duas usinas nucleares e vislumbra a construção da terceira. Além das usinas, com suas estruturas operacionais, temos também todas as áreas de suporte administrativo da empresa. Portanto, há na empresa um grande número de usuários em potencial para utilizar a tecnologia de mineração de textos. Cada usuário ou comunidade interna de usuários tem suas necessidades e interesses específicos, lidando também com diferentes tipos de documentos. Na ELETRONUCLEAR há bases internas de documentos de natureza administrativa e natureza operacional – a experiência operacional.

No que tange à Experiência Operacional, muitas bases de documentos se encontram hoje disponíveis, especialmente após a implantação da Gestão Eletrônica de Documentos e do Portal Corporativo. Há bases internas e bases externas acessadas sob contratos, convênios, acordos ou aquisição do direito de acesso.

A Experiência Operacional Externa é constituída por bases englobando um grande número de documentos oriundos de instâncias como:

- AIEA – Agência Internacional de Energia Atômica
- CNEN - Comissão Nacional de Energia Nuclear

- WNA – Word Nuclear Association (Associação Nuclear Mundial)
- WANO (*Word Association of Nuclear Operators*)
- INPO (*Institute of Nuclear Power Operators*).
- NRC – *United States Nuclear Regulatory Commission*
- VGB – Instituição européia, sediada em Essen, na Alemanha e que atua na área nuclear.
- Instituições Nacionais de Pesquisa
- Instituições Estrangeiras de Pesquisa
- Fornecedores
- Outras empresas da área de energia nuclear.

No que concerne à Experiência Operacional Interna, ela encontra-se subdivida por usina. Há que se considerar o fato de que estão em operação duas usinas nucleares para geração de energia, construídas em momentos diferentes a partir de fornecedores diferentes. Isso implica que temos estruturas gerenciais, processos e bases de documentos separadas para cada uma das duas usinas. Conforme apresentado no capítulo 9, Angra I foi construída com tecnologia da empresa Westinghouse® (EUA) e iniciou sua operação em 1985. Já Angra II, iniciou sua operação em 2000 e foi construída com tecnologia da empresa alemã Kraftwerk Union A.G. – KWU, então subsidiária da Siemens®. Há ainda o projeto de construção de Angra III. Portanto, hoje, há bases consolidadas de documentos relativos à Experiência Operacional separadas para as Usinas de Angra I e Angra II.

Diante dos aspectos acima relacionados, no âmbito do presente trabalho, considerou-se necessário uma pesquisa com duas vertentes:

- Estudar aspectos relevantes concernentes à Experiência Operacional da ELETRONUCLEAR que deveriam ter prioridade para investigação e análise, para implantação de um processo de descoberta de conhecimento em bases textuais (mineração de textos), especialmente no contexto do projeto Portal-GEDT-*Workflow*, anteriormente implantado.

- Identificar, adequar e aplicar o processo de descoberta de conhecimento em bases textuais com o *software* de mineração de textos em um conjunto de experimentos, avaliando os resultados dessa aplicação quanto à relevância dos resultados e grau de aderência às necessidades dos usuários e da própria organização.

Para a primeira vertente, estabeleceu-se a condução uma pesquisa-ação para responder às questões nela colocadas e, em seguida, para a segunda vertente, deveríamos aplicar software de mineração de textos em um projeto de experimentos com bases documentais selecionadas pela empresa.

11.3 A Pesquisa-Ação na Eletronuclear

11.3.1 Visão Geral e Escolha da Questão-Alvo do Processo de Mineração de Textos

A Pesquisa-Ação é um método ou abordagem qualitativa de pesquisa, conforme discutido no capítulo 3 da presente Tese. Portanto, o primeiro desafio do autor dessa Tese foi entender a ELETRONUCLEAR, o setor nuclear, suas características, história, situação atual, perspectivas futuras, etc., bem como aprofundar a percepção sobre a adequação do uso da tecnologia de mineração de textos na empresa. Em seguida tornou-se necessário estudar e entender os aspectos relativos ao conhecimento – tácito e explícito da Empresa e aprofundar o conhecimento das necessidades de informação existentes junto às diferentes comunidades de profissionais em diferentes áreas, para inferir respostas que seriam adequadas.

A relação principal do autor da Tese com a ELETRONUCLEAR deu-se com a Gerência de Engenharia de Apoio, que tem por missão fornecer o apoio de Engenharia necessário à operação, visando maior segurança e melhor desempenho das usinas de Angra I e II. A seguir, apresentamos a Figura 48 com uma visão geral da GEA.T.

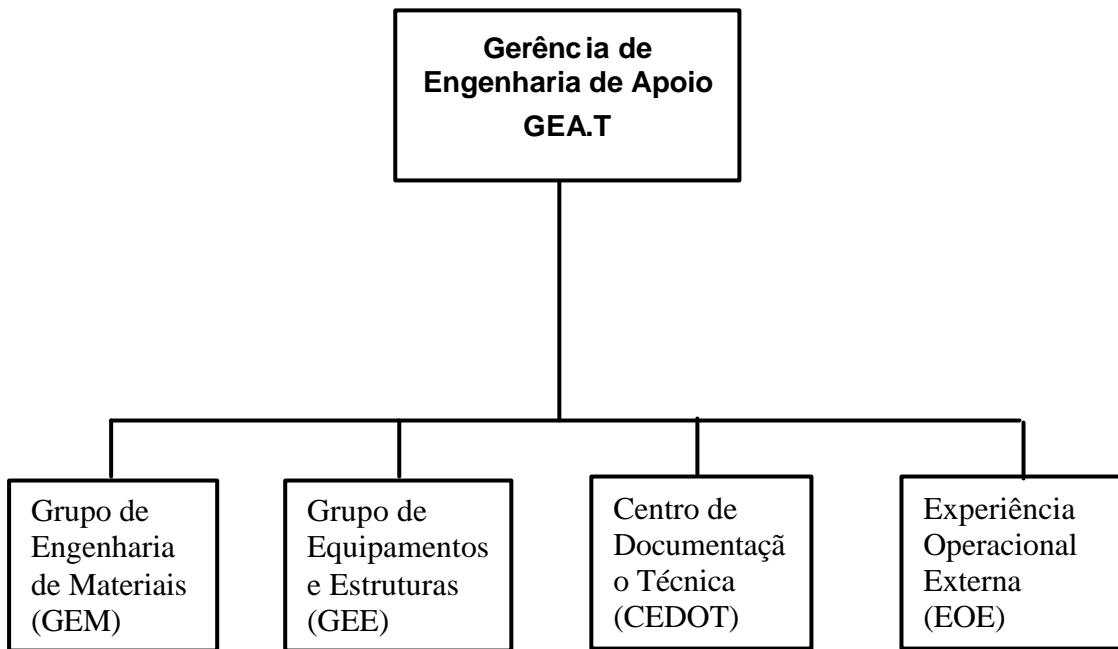



Figura 48: Estrutura Geral da GEA.T
Fonte: ELETRONUCLEAR.

A Gerência de Engenharia de Apoio – GEA.T está subordinada à Superintendência de Engenharia de Apoio, a qual por sua vez, está vinculada à Diretoria Técnica da ELETRONUCLEAR. A GEA.T atua principalmente no apoio de engenharia às usinas Angra 1 e Angra 2, com as seguintes atribuições:

- Prestar apoio de engenharia às usinas abrangendo Sistemas, Estruturas, Componentes e Confiabilidade de Materiais.
- Prestar suporte imediato de engenharia às usinas em operação.
- Gerenciar o Centro de Documentação Técnica (CEDOT) abrangendo os arquivos de originais de Engenharia de Angra 1 e 2, e os arquivos de registros da qualidade e modificações de projeto de Angra 1, 2 e 3.
- Coordenar o Gerenciamento Eletrônico da Documentação de Engenharia.
- Gerenciar os processos de acompanhamento, avaliação e disseminação e implementação da Experiência Operacional Externa.
- Prestar serviços de inspeções periódicas, instrumentação e reparos de estruturas civis das usinas em operação.

- Prover suporte técnico às usinas em atendimento a solicitações de Esclarecimentos Técnicos da Operação (ETO).
- Prestar apoio de engenharia às usinas durante as paradas e nas inspeções em serviço.
- Prestar serviços em engenharia de componentes e em projeto nas áreas mecânica, elétrica e civil.
- Desenvolver e coordenar programas de longo prazo tais como o Programa de Gerenciamento da Obsolescência, o Programa de Liga 600, o Programa de Controle da Corrosão por Ácido Bórico, o Programa de Dedicação e o Programa de Monitoramento de Mecanismos de Degradação.
- Prestar apoio técnico às usinas na elaboração do orçamento de investimento e na avaliação de propostas técnicas e de contratos.
- Planejar atividades para o Plano de Metas da DT.
- Prestar apoio na elaboração dos orçamentos de investimento e estrutural.

É importante salientar que um aspecto que atravessa todas as discussões sobre aplicação da mineração de textos em bases de documentos é justamente o valor e a utilização do conhecimento tácito dentro da organização, conhecimento esse embutido nas mentes e na expertise dos profissionais da área nuclear e, no caso em questão, da ELETRONUCLEAR. O conhecimento tácito pode migrar, com alguma facilidade, para os textos, pela flexibilidade e facilidade de elaboração que este possui hoje, já em formatos digitais, ou seja: a elaboração de textos pode ser uma forma de conversão de parcelas de conhecimento tácito em conhecimento explicito. Conforme RINTALA & KURONEN (2006) o conhecimento tácito é altamente subjetivo e dependente de um número grande de fatores intangíveis, tais como cultura, opiniões pessoais, valores, perspectivas, *insights*, intuições e emoções do indivíduo. Ele é construído através do esforço para construção de um sentido de mundo. O conhecimento tácito é individual e subjetivo. De acordo com esses autores, há diferentes visões sobre como o conhecimento tácito pode ser convertido em conhecimento explicito. Há pesquisadores que afirmam que esse conhecimento nem pode ser convertido (JOHANNESSEN et alli 2001)mas pode somente ser compartilhado em interações sociais entre os profissionais

experientes e os iniciantes através de uma vivência e aprendizagem quase artesanal destes com aqueles. Outros pesquisadores acham que o conhecimento pode ser parcialmente codificado. Isso pode ser feito em um nível conceitual, mas não com grande detalhamento. Por fim, há também os pesquisadores que afirmam que esse conhecimento pode e deve ser compreendido e codificado para compartilhamento. A codificação pode ser alcançada através de alguns caminhos, tais como codificação de conhecimento em fórmulas, códigos, relatórios, políticas e incorporado em objetos físicos como os protótipos.

Essa codificação do conhecimento tácito tem claras vantagens: mais itens de conhecimento podem ser codificados, gerando mais economia e eficiência e pode ser compartilhado. Entretanto, essa codificação pode criar novos custos para a organização, como com tecnologia da informação e documentação excessiva. É necessário refletir, portanto, sobre a parcela de conhecimento tácito que deve ser explicitado e codificado de forma equilibrada. Estudos prévios indicam que essa proporção pode variar de uma organização para outra.

RINTALA & KURONEN (2006) propõe a utilização de 23 métodos de compartilhamento de conhecimento tácito e que podem ser adotados pela ELETRONUCLEAR. Tais métodos são:

1) Aprendizagem Individual Com Especialista – provavelmente o mais antigo método de partilhar conhecimento tácito e que envolve a consultoria assistida de um profissional experiente para guiar o aprendiz em situações da vida real, sendo que este pode observar gradualmente a responsabilidade na performance de tarefas de trabalho.
2) Aprendizagem em Grupo com Especialistas
3) Mentoring individual
4) Mentoring em Grupo
5) Visitas técnicas a locais de operação e produção, com acompanhamento de especialistas
6) Observação
7) Discussão de Casos

8) Implantação de salas de jogos, utilizando conversas informais, uso de jogos e processos lúdicos de aprendizagem, considerando a premissa de que o relaxamento dos participantes leva a uma maior socialização do conhecimento.
9) Reuniões em grupos multifuncionais, com realização de cruzamentos das experiências de diferentes matizes e áreas que podem se complementar.
10) Reuniões em grupos de especialistas com análise em profundidade de problemas e questões específicas
11) Trabalho em conjunto de novatos e "*experts*" em um time com oportunidade de observar e conceituar situações, abordar problemas, generalizar e avaliar soluções.
12) Revisão de rotinas organizacionais existentes com desenvolvimento de novos modelos para novas situações a partir de discussões em grupos de questões diversas relativas aos processos organizacionais existentes.
13) Grupos de aprendizagem focada, em que aprendizes formalmente sustentam posições de aprendizagem, participando de reuniões em trocam contribuições com seus colegas relativas aos estudos e atividades desenvolvidas.
14) Comunidades de prática, processo de interação em grupo em que comunidades emergem informalmente, construindo redes de conhecimento de especialistas e podem envolver pessoas de dentro e de fora da organização no sentido de resolver problemas complexos para os quais soluções desafiadoras são necessárias, porém, difíceis de serem individualmente construídas. São constituídos usando reuniões, trocas de mensagens eletrônicas, uso de videoconferências sobre tópicos de interesse de seus membros.
15) Jogo de simulação em que especialistas jogam e aprendizes observam. Os observadores colocam situações problemáticas de trabalho. Durante o processo de solução, os especialistas colocam suas experiências em palavras e os novatos podem aprender sobre aspectos relativos à tomada de decisão. Após os participantes terem colocado suas visões, desenvolvem em conjunto a solução do problema.
16) Relatos de histórias em que narrativas são apresentadas, mostrando um contexto, a trajetória de situações, eventos e problemas e quais os desfechos ocorridos.
17) Grupos de relatos de histórias. Neles, mais pessoas são envolvidas no processo e contribuem na criação de histórias, interagindo com opiniões, visões e

interpretações de eventos passados. Esses grupos exigem que especialistas invistam boa parte do seu tempo na estruturação e composição das histórias a serem discutidas.

18) Uso de entrevistas estruturadas
19) Uso de entrevistas semi-estruturadas
20) Realização de estudos sistematizados, com levantamento e agrupamento de dados sobre um determinado tema para posterior análise, visualização das abordagens existentes sobre determinada questão ou fenômeno e estabelecimento de teorias ou pressupostos sobre ela e suas possíveis abordagens.
21) Escrita e redesenho de padrões
22) Uso de árvores de decisão
23) Implantação de times de melhores práticas.

Podemos observar que o método proposto no item 20 se alinha de modo forte com a tecnologia de mineração de textos.

Para realização da Pesquisa-Ação na ELETRONUCLEAR, visando entender, avaliar e introduzir o uso da tecnologia de mineração de textos na organização, foi constituído um grupo focal, de acordo com o descrito no tópico 3.3 da presente tese com seis integrantes para participação do processo, sob condução do autor da Tese. Tal grupo incluiu, além do autor, dois componentes da equipe de desenvolvimento do software ***DifBrain***®[97], e três profissionais[98] da Gerência de Engenharia de Apoio – EOE – Experiência Operacional Externa, área subordinada à Diretoria Técnica da ELETRONUCLEAR e que foi a responsável pelo Projeto Portal GEDT *Workflow* , sendo hoje a área executiva responsável pela articulação dos processos relativos à Gestão do Conhecimento na Empresa. Portanto, em análise conjunta com a Gerência de Engenharia de Apoio procurou-se avaliar em que campo do conhecimento sobre energia

[97] Analista de Sistemas Marcelo Gibson de Castro Gonçalves e a Professora Doutora Regina Fátima Gibson Gonçalves, Engenheira Eletricista, especialista na área de Inteligência Artificial.

[98] O Sr. Luiz Celso Oliveira, *Coordenador de Experiência Operacional Externa - EOE* e já citado nessa Tese, e os Srs. Carlos Maurício Cruz Belo e Ricardo Lowe Stukart, Engenheiros de Experiência Operacional. Nesta função executam análise da experiência operacional externa, e dão suporte de engenharia às usinas na análise de experiência operacional interna. Esses profissionais são experientes no campo da Engenharia Nuclear, possuidores de competência no campo na análise de documentos da Experiência Operacional das duas usinas em operação em Angra dos Reis.

nuclear seria mais adequado desenvolver o processo de mineração de textos, numa perspectiva de envolvimento com a busca de solução para problemas críticos e necessidades relevantes que a ELETRONUCLEAR atualmente possui. Sendo interesse da Empresa o uso da tecnologia de mineração de textos, procurou-se identificar em quais aspectos da realidade organizacional seria possível constatar maior aderência entre as necessidades de potenciais usuários e o uso da tecnologia proposta.

Embora em várias áreas da ELETRONUCLEAR, em conformidade com as perspectivas descritas na literatura, seja possível visualizar várias possibilidades de aplicação, tornou-se necessário delimitar o foco de aplicação da mineração de textos, pelo menos em sua fase inicial e experimental. O foco recaiu sobre a Experiência Operacional, que possui um considerável acervo de documentos digitais. Em seguida, optou-se também por tratar as bases de documentos da Experiência Operacional Interna das usinas de Angra I e Angra II. Um extrato da *Guideline* da *WANO (World Association of Nuclear Operators)* usada como resumo de referência básica para atividades de gestão da Experiência Operacional é apresentada no ANEXO IV da presente Tese.

11.3.2 Análise de Eventos em Suporte ao Monitoramento de Usinas Nucleares

O grupo constituído examinou diferentes possibilidades de problemas que poderiam ser abordados, todos envolvendo a necessidade de descoberta de conhecimento embutido em bases de documentos da organização. A "Análise de Eventos em Suporte ao Monitoramento de Usinas Nucleares" foi o problema escolhido para estudo e entendimento pelo grupo e para posterior aplicação da tecnologia de mineração de textos. Essa é uma questão fortemente inserida no âmbito da Experiência Operacional, tanto Externa quanto Interna, embora para o uso da ferramenta de mineração de textos, em um primeiro momento e para o conjunto de experimentos realizados, foram utilizadas apenas bases de documentos digitais da Experiência Operacional Interna.

A Experiência Operacional constitui um conjunto de conhecimentos de alta relevância sob o ponto de vista gerencial das usinas nucleares. Ela configura um contexto básico e um parâmetro fundamental para a tomada de decisões relacionadas à

gestão de operação. A gestão da operação deve buscar aderência e alinhamento entre ações de otimização de desempenho e zelo pela segurança. A intenção da empresa é sempre procurar gerar o máximo possível de energia com as plantas em operação, mas isso deve ser alcançado observando sempre as melhores práticas de operação e segurança. Gerenciar a operação das usinas implica em um processo de monitoramento que demanda conhecimento sobre o status, a condição da usina no presente, mas a interpretação de tais estados pode ser sempre subsidiada pela história passada, pela análise e interpretação de eventos anteriores e que foram fatores de aprendizagem para os gestores e empregados responsáveis pela operação. Gerenciar uma usina nucleoelétrica de modo a garantir uma operação otimizada e segura é o ideal a ser buscado, especialmente quando se considera a complexidade do processo, pela quantidade de fatores intervenientes envolvidos. Conciliar segurança e produção é hoje algo imprescindível e o acúmulo da experiência operacional vem contribuindo para que se possa melhor entender os parâmetros relativos ao que seja uma boa performance e uma adequada segurança para operação de usinas nucleares. Atualmente, identificar, monitorar, mensurar e avaliar esses parâmetros tornou-se um imperativo para todos os operadores de usinas ao redor do mundo, o que levou entidades relacionadas à área nuclear a incorporar tal preocupação e desenvolvessem a tarefa de especificar referenciais para a operação, avaliando níveis de segurança e desempenho das plantas nucleares. Destarte, instâncias como AIEA, INPO, WANO e NRC vêem desenvolvendo trabalhos concernentes à definição de indicadores de desempenho e segurança.

De acordo com SOUTO (2005) a AIEA, desenvolveu o *Operational safety performance indicators for nuclear power plants* (IAEA, 2000)[99], que apresenta um guia geral de indicadores de desempenho e segurança. Tal guia é de relevante importância pelo fato de estabelecer uma nova visão no que tange ao tratamento de aspectos relacionados a desempenho e segurança de usinas nucleares. Ele sugere ações, condições e procedimentos para se atingir os requisitos de segurança, bem como garantir um alto índice de desempenho, propondo que problemas relacionados a desempenho e segurança sejam tratados de forma macro e não simplesmente através de indicadores específicos isoladamente. A Agência afirma que um alto nível de segurança

[99] IAEA 2000, "*Operational Safety Performance Indicators for Nuclear Power Plants*" TECDOC-1141, www.iaea.org,. Citado por SOUTO (2005).

é resultado de uma interação complexa de uma segurança operacional, projeto e desempenho humano. O autor afirma:

> *"A Experiência tem demonstrado que a concentração em um único aspecto do desempenho é ineficiente e pode ser ilusório. O que é mais válido, portanto, é a figura total apresentada por um conjunto completo de indicadores projetado para monitorar todos os aspectos do desempenho de segurança operacional. Destarte, tal guia geral de indicadores fornece uma estrutura genérica para a identificação e organização dos indicadores de desempenho que têm uma relação direta com os atributos de segurança desejados. O trabalho da Agência mostra ainda que as tendências de um indicador específico em um período de tempo podem fornecer um aviso para que se possam investigar as causas por trás das alterações observadas. Além é claro de servir como instrumento de comparação dos indicadores com as metas e objetivos identificados para avaliar pontos positivos e negativos de desempenho* (p. 2-3)".

SOUTO (2005) também comenta sobre outros órgãos e institutos e descreve que o INPO também traz alguns trabalhos relacionados aos indicadores de desempenho, dos quais merece destaque o *Indicators of changing performance* (INPO, 2001)[100], destacando a importância de se estabelecer uma estruturação eficiente de indicadores de desempenho e segurança operacional, para monitorar aspectos da planta, organizacional e de desempenho humano, além de identificar um conjunto valioso desses indicadores. Já a WANO, de acordo com esse autor, desenvolve um programa de padronização mundial dos indicadores de desempenho de usinas nucleares, enquanto que a NRC possui larga experiência sob indicadores de desempenho, resultado da implementação do primeiro programa oficial desses índices em usinas nucleares (NRC, 2002)[101].

A análise de eventos é uma atividade de relevância no sentido de melhorar os indicadores e subsidiar ações de melhoria para segurança e desempenho. A análise de eventos, realizada com o suporte da tecnologia de mineração de textos, permite a emergência de certos conhecimentos que são fruto do tratamento inteligente realizado por ferramenta de *software* sobre esses mesmos textos. Conforme apresentado no capitulo 10, as notificações de eventos ocorrem com base na relevância do mesmo, classes de significância e suas conseqüências, grau de segurança, prazos de

---

[100] INPO 2001, 01-005, *"Indicators of Changing Performanc*e", www.eh.doe.gov/inpo, Citado por SOUTO (2005).

[101] NRC 2002, 0030, *"Summary Report on NRC's Historical Effots to Develop and Use Performance Indicators"*, www.nrc.gov/reading-rm/doc-collections/comission/secys/2002/secy 2002-0030/2002-0030.html Citado por SOUTO (2005).

comunicação, dentre outros aspectos. Tais relatórios são gerados, sob a forma de documentos textuais eletrônicos e comunicados às instâncias competentes, internas ou externas à empresa, conforme o caso. A Eletronuclear tem hoje uma base significativa dos seguintes relatórios[102], cada um com sua respectiva estrutura, tanto para Angra I quanto para Angra II:

- Relatórios de Evento – RE
- Relatórios de Desvio Operacional – RDO
- Relatórios de Evento Significante – RES
- Relatórios ORO – Ocorrência de Relato Obrigatório

Na medida em que esses documentos pudessem não somente ser alvo de uma boa recuperação por parte dos diversos usuários (o que foi alcançado com o uso do Portal de Experiência Operacional), mas também se transformassem em alvo de um processo de análise e de um tratamento computacional que rastreasse e identificasse o seu real conteúdo, realizando-se sobre eles uma análise semântica e sinóptica[103] eles poderiam se transformar em fonte valiosa para profissionais envolvidos na operação das usinas no sentido de identificar e visualizar múltiplos fatores intervenientes em *performance* e segurança.

11.4 O Projeto de Experimentos na Eletronuclear

11.4.1 Proposta do Projeto

Partindo das necessidades de informação e de conhecimento da Eletronuclear, já caracterizadas, a partir do seu projeto de Gestão do Conhecimento, também já apresentado e que indicou a pertinência e adequação do uso da tecnologia de mineração

---

[102] Conteúdo e desenho desses relatórios não serão aqui apresentados por força de acordo de não divulgação de informações relevantes e estratégicas da ELETRONUCLEAR pelo autor dessa Tese. Esses relatórios envolvem informações sobre ocorrências diversas em várias áreas da usina: envolvendo a operação do reator, gerador, processo de resfriamento nos diferentes circuitos primário e secundário, dentre outros processos e sub-processos operacionais. Ver também no ANEXO II sobre relatórios.

[103] Por análise sinóptica entende-se um olhar único sobre um mesmo conjunto. Um acervo de documentos possui muitos itens. Ao recuperar esses itens, o usuário procede uma análise individual de cada item, mas tem dificuldade em estabelecer correlações entre os documentos e entre os diferentes conteúdos possíveis de existir em cada um dos documentos. A análise sinóptica implica em ver a totalidade, encontrando significado ou sentido semântico não apenas em um documento específico, mas no conjunto objeto de análise.

de textos já discutida, foi definido um projeto de experimentos para uso dessa tecnologia no âmbito da organização, visando abordar a Análise de Eventos em Suporte ao Monitoramento de Usinas Nucleares, conforme acima citado.

O projeto de experimentos assumiu os seguintes objetivos:

a) Identificação de necessidades de informações específicas de subgrupos de usuários, funcionários da Eletronuclear, com demanda de informação relevante que se encontrava inserida em documentos textuais, no caso em questão para abordar o problema: Análise de Eventos em Suporte ao Monitoramento de Usinas Nucleares;
b) Determinação de bases dados textuais de interesse de um determinado subgrupo de usuários que deveriam conter as informações textuais relevantes a serem recuperadas;
c) Determinação dos tipos de análises e conhecimento a serem descobertos que deveriam ser alvo do processamento dos textos pelo software ***DifBrain***®
d) Uso do software ***DifBrain***®, na recuperação da informação textual nas funções propostas para o mesmo, com o processamento dos textos em computador;
e) Avaliação dos resultados em conjunto com o usuário.

No que tange aos tipos de análises e conhecimento a serem descobertos que deveriam ser alvo do processamento dos textos, foi estabelecido pelo Grupo Focal os seguintes objetivos finais para a Mineração de Textos:

- Descobrir conhecimento embutido nos textos dos relatórios citados e que não estavam evidentes ou claros para uma análise tradicional ou busca convencional;
- Realizar buscas proativas e reativas (conforme descrito no tópico 6.4.4) de forma a identificar possíveis relações expressivas de conhecimento relevante a ser apresentado após a execução das tarefas de mineração (conforme descrito no item 8.7).
- Identificar possíveis tendências nos relacionamentos de conceitos e termos presentes nos contextos dos documentos.

- Promover o tratamento semântico de coleções em que havia ambigüidade de palavras presentes em relatórios.

O software DifBrain® começou então a ser utilizado no processamento das bases textuais da ELETRONUCLEAR.

11.4.2 Execução dos Experimentos

Entre 2004 e 2007, foi desenvolvido o *software* ***DifBrain***®, já caracterizado, pela Diferencial Consultores Associados®. O software ***DifBrain***®:

- Gerencia e processa bases de textos com técnicas de mineração de dados
- Opera em 3 idiomas (português, inglês e alemão)
- Gerencia de idiomas, eliminando stopwords, aplicando delimitadores de sentenças, alfabeto especial e sinais de pontuação
- Implementa buscas pro-ativas
- Implementa buscas reativas
- Gera resumos automáticos
- Cria clusters de documentos, implementando diferentes algoritmos
- Gera estatísticas com freqüência de termos em bases de documentos processados

Em cumprimento a tal especificação do tópico anterior, foram identificadas as bases textuais armazenadas no computador servidor *Natrium* da Eletronuclear, dedicado à Experiência Operacional Interna[104]. Elas incluem os documentos RE, RDO, RES e ORO para as usinas de Angra I e II. Esses documentos se encontravam no formato .doc e .pdf. O interesse imediato de estudo dessas bases foi a análise sinótica – estudo dos relacionamentos no contexto dos documentos, com a pergunta do que poderia ser descoberto com múltiplos cruzamentos.

Inicialmente foi realizada a etapa de pré-processamento, conversão das bases para tipos de documentos comuns, visando facilitar e permitir a atividade posterior de mineração. Assim, todos os documentos foram convertidos ou para arquivos .pdf, .rtf ou

[104] O usuário poderia escolher a base de textos que desejasse.

.txt – formatos mais adequados à atividade proposta para o software ***DifBrain***®. Figura 49

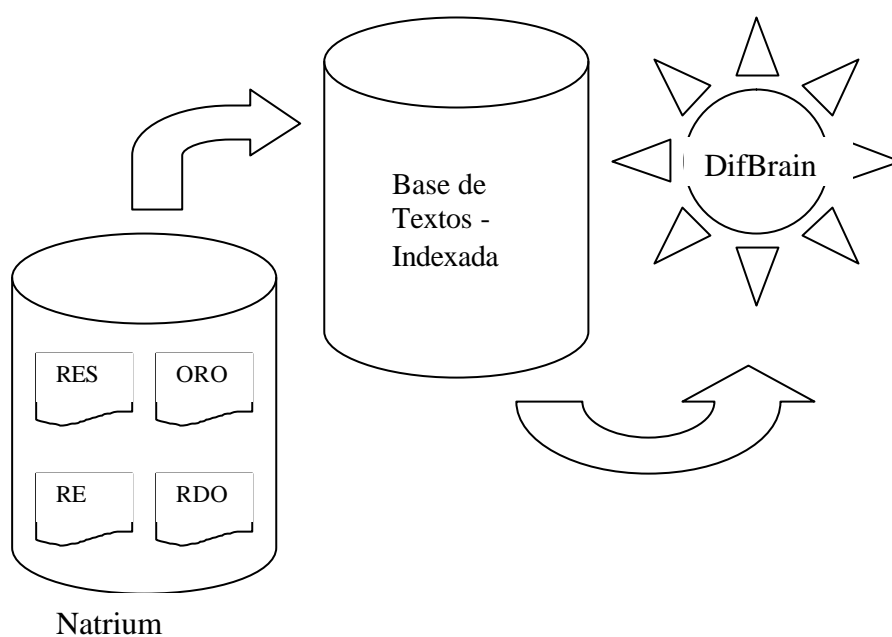



Figura 49: Representação do Processamento – Software de Mineração de Textos

Em um segundo momento, houve a etapa de indexação automática, ou seja: o software ***DifBrain***® criou as representações dos documentos a serem tratados em suas próprias bases, basicamente compostas de documentos .txt para facilitar o processamento ulterior. A base utilizada para o processo experimental continha 2000 documentos, sendo cerca de 2/3 relacionados à usina de Angra I, que é mais antiga, e o restante de Angra II.

Em seguida o software ***DifBrain***® realizou as tarefas de:

- Categorização
- Agrupamento (ou *Clustering*)
- Geração de Resumos Automáticos
- Tratamento Estatístico do Documento

Os resultados dessas tarefas é que são examinados para avaliação da capacidade da mesma na descoberta de conhecimento em bases textuais. Aconteceram dezenas de "rodadas de processamento" sobre as bases textuais da ELETRONUCLEAR – cada uma delas podendo ser considerada um experimento, até a identificação dos resultados satisfatórios, o que foi obtido pela adequação da ferramenta (customização) às necessidades e desafios da organização pesquisada. O Grupo Focal sugeriu a mesclagem de critérios para elaboração de clusters e categorização de documentos. Assim, por exemplo, poderíamos ter categorizações e agrupamentos que levassem em conta o cruzamento de critérios, tais como: estudar eventos dos RESs e RDOs com base na interseção das visões de potencia do reator, data e ocorrência de eventos semelhantes. Isso significou buscar nos documentos processados visões como: que tipos de eventos com alguma similaridade acontecem quando o reator opera em uma determinada potencia?

Outras perguntas sobre os resultados e a atuação do software poderiam ser:

- Quais as principais falhas humanas estão acontecendo quando o gerador opera a uma determinada capacidade?
- Quais as correções de falhas foram recomendadas em 2005 e que ocorreram de novo em 2006, quando o reator operava a uma potencia de x% ?
- Qual a causa raiz latente de um evento cujo relatório indica uma coisa em seu cabeçalho e outra em seu corpo de texto?

Para todos os documentos processados, foram definidos parâmetros – linhas orientadoras para – o processo de descoberta nos textos, tais como: potência do reator, potencia do gerador, modo de operação, suspeita de falha humana, correção de falha humana recomendada, período (dia, mês, ano, etc). Isso implica que o usuário pode escolher qual desses critérios ele quer que seja levado em conta para uma análise sinóptica do acervo de documentos que lhe interessa. Isso é indicado nas Figuras 50 e 51 abaixo.

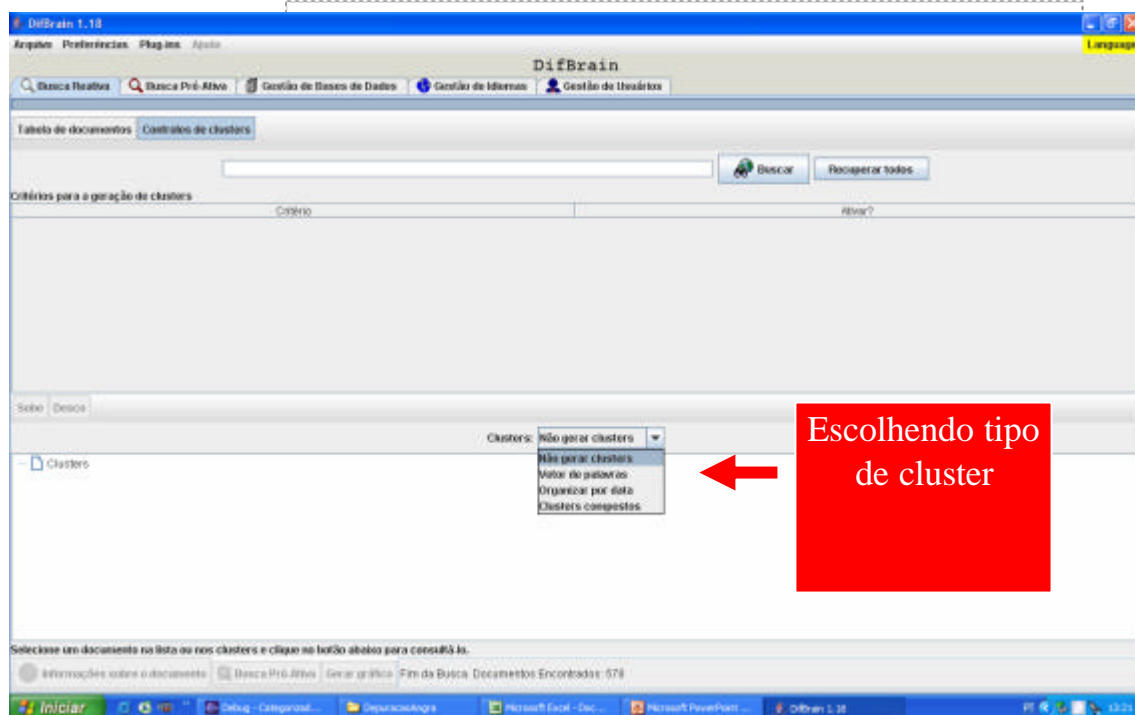



Figura 50: Tela DifBrain – Escolhendo os tipos de cluster a ser utilizado no processamento (a).

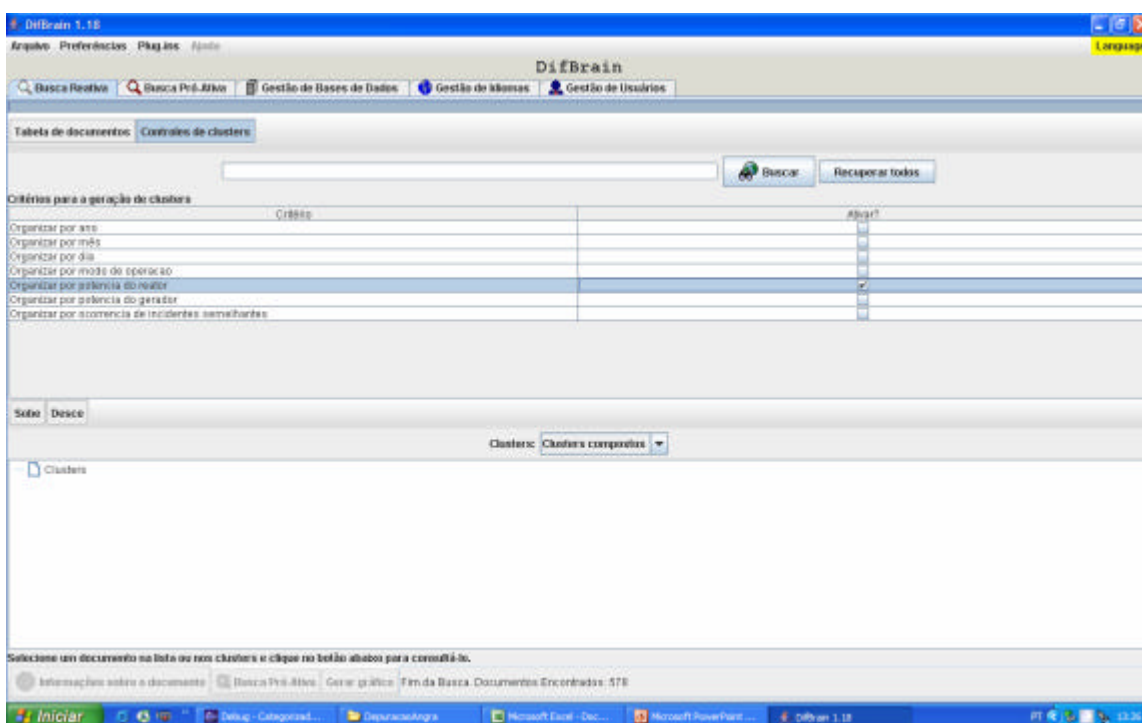



Figura 51: Tela DifBrain – Escolhendo os tipos de cluster a ser utilizado no processamento (b).

Os textos são processados e há geração do resumo automático, feita a *clusterização* e a categorização. A figura 52 apresenta a tela de geração de resumo automático.

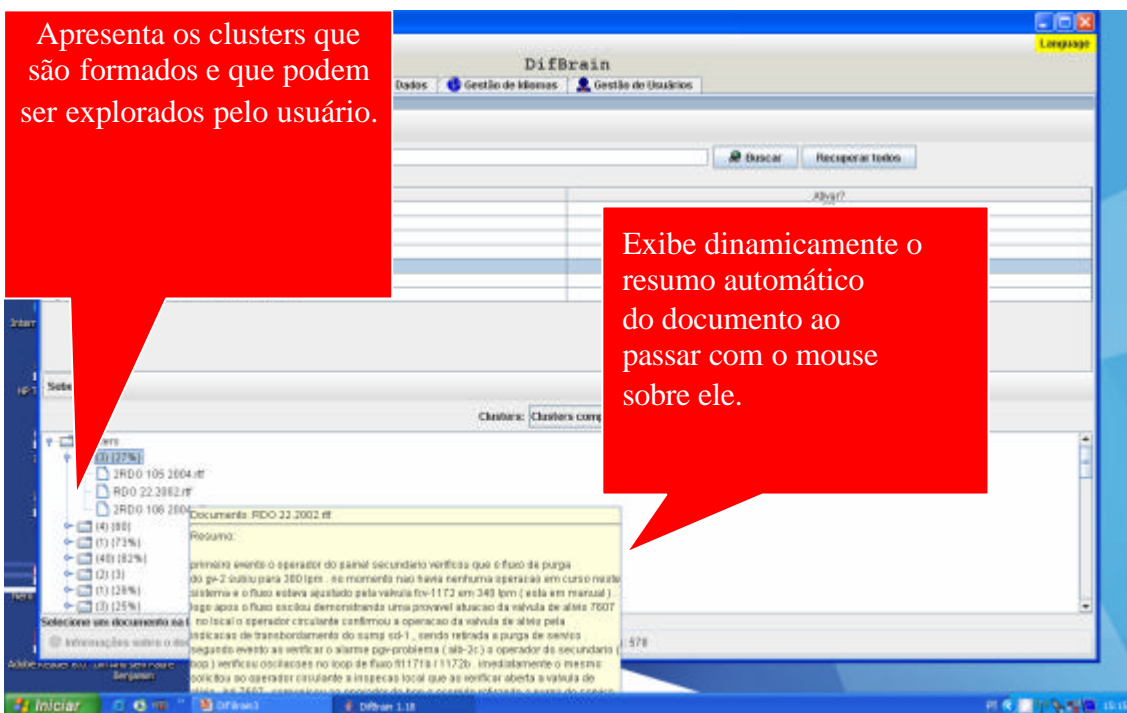



Figura 52: DifBrain gera resumo automático de um documento.

11.4.3 Avaliação dos Experimentos

A avaliação dos experimentos levou, sobretudo em conta, as tarefas de descoberta de conhecimento em bases de textos realizadas pelo software ***DifBrain***®. No processo de avaliação, os primeiros aspectos considerados foram a seleção de bases de dados – considerada fácil e interativa, portanto, adequada para o usuário e a geração do resumo. A geração automática do resumo é um processo sujeito a um constante aprimoramento. Ele leva em conta aspectos semânticos (significado e relevância de termos e frases no documento) e estatísticos (freqüência desses elementos no documento). Pode levar também em conta a estrutura do documento ou mesmo se já existe algum resumo ou síntese preliminar. Os resumos automáticos gerados foram considerados pelos profissionais da ELETRONUCLEAR como adequados, portando significados que expressam o real sentido do documento.

Um segundo aspecto avaliado foi a geração de clusters dos documentos. ***DifBrain***® mostrou-se uma ferramenta poderosa não só por *clusterizar*, mas por

agregar diferentes visões nos clusters e realizar a mesclagem (combinação) de diferentes critérios para sua geração. Assim é possível construir uma matriz de combinação desses diferentes critérios, cujos resultados para o usuário foram considerados expressivos. As visões passíveis de extração dos clusters representam conhecimento novo, antes não percebido numa determinada perspectiva. Por exemplo, como comprovar que a incidência de certos eventos aumenta ou diminui com as variações de potência do gerador ou do reator? A partir dessa descoberta, que ações corretivas podem ser propostas? E a partir dessa constatação, ficou comprovado que é possível fazer outras e sucessivas combinações, como por exemplo, com suspeita de falha humana. A ferramenta cumpriu o papel de realizar também pesquisas sucessivas dentro dos clusters gerados, seja por tipo de evento ou por palavra.

No que tange à categorização, foi também constatada sua adequação, demandando apenas alguma melhora na visualização, uma vez que as categorias aparecem na mesma estrutura do cluster, o que pode confundir um pouco o usuário. Entretanto, seus resultados são bons, como demonstra o breve caso narrado a seguir.

11.4.4 Mineração de Textos Aplicada ao Problema da Analise de Falhas Humanas.

De acordo com estimativa dos profissionais da ELETRONUCLEAR envolvidos no presente trabalho, cerca de 60% das falhas relatadas nos relatórios de desvio operacional e nos relatórios de evento significante são causados por falha humana. Dentro da idéia de monitoramento e de uma ação gerencial que visa conciliar desempenho e segurança, tal fato precisa ser minimizado. E somente o será se suas causas forem devidamente exploradas. É necessário sempre estudar os eventos de modo a se chegar a uma compreensão efetiva de sua causa-raiz.[105] Isso vale para todos os eventos, mas tem ainda maior relevância na análise de falhas humanas.

Em muitos relatórios internos – RES, RDO e ORO, não são explicitadas como causa de falha as falhas humanas. Os motivos são, sobretudo, técnicos ou de outra ordem. Oculta sobre um motivo técnico, entretanto, encontra-se uma falha humana. Muitas vezes, portanto, para evitar uma explicitação forte, a causa técnica é explicita,

---

[105] ***Causa-raiz*** - É a causa fundamental de um determinado *evento* que, se eliminada, evitará a sua ocorrência ou repetição, ou seja: se não tivesse acontecido isso, não teria acontecido aquilo.

mas a humana não aparece objetivamente. Entretanto, analisando os textos dos documentos, percebe-se que há uma falha humana subjacente. A detecção desse fenômeno é complexa e pressupõe, do ponto de vista humano, a leitura minuciosa e individual de cada documento. Para tratar essa questão, a ferramenta ***DifBrain***® mostrou-se adequada, por elucidar aspectos ocultos desses relatórios. Isso foi conseguido com o uso da Indexação Semântica Latente (descrita no capitulo 5). Na ISL temos um espaço conceitual foco de atenção e análise. Temos também os termos que representam esse conceito ou evento de interesse a ser buscado, mas muitas vezes ele não aparece nas consultas, pois os usuários usam termos diferentes para falar a mesma coisa. Palavras e frases podem estar nos documentos de modo muito diferente dos termos representativos, embora estejam dentro do mesmo espaço conceitual Figura 21 abaixo, já apresentada no capitulo 5.

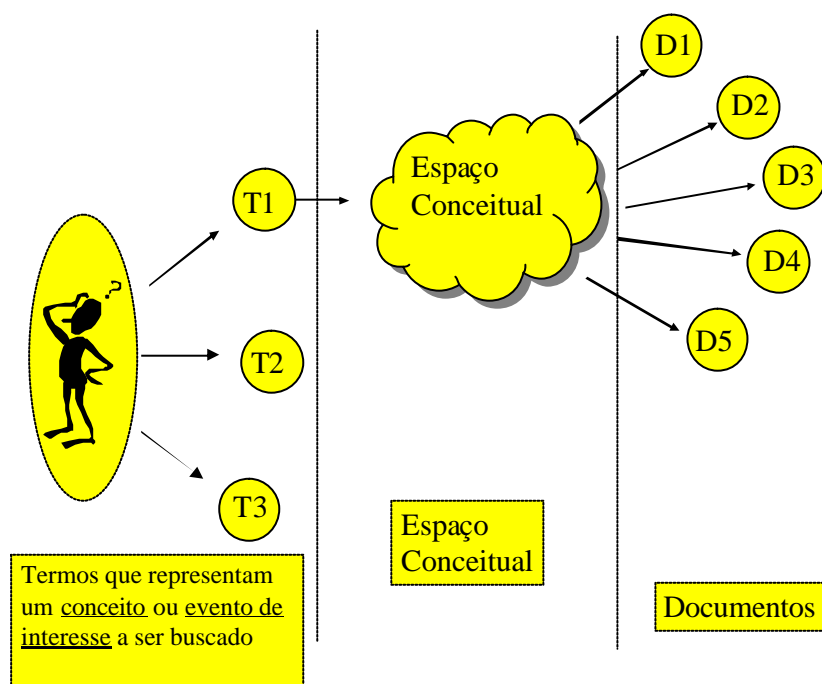



Figura 21: Uma representação da Indexação Semântica Latente

No caso da análise de falhas humanas, esse processo foi aplicado sobre os relatórios da ELETRONUCLEAR. Eles foram tratados pela ferramenta de mineração de textos e categorizados sob três grupos:

- Documentos que diziam não ter ocorrido falha humana, mas que possuíam realmente falha humana embutida;
- Documentos que diziam ter falha humana;
- Documentos de diziam não ter ocorrido falha humana e realmente não ocorreu.

Através do uso da ISL, foi possível rastrear e analisar frases e expressões indicativas da presença de tais falhas para o primeiro grupo de documentos. A figura 53 abaixo mostra o processo, indicando possíveis frases nos documentos (à direita) que indicaram essa presença no processamento dos textos. Concluindo: independentemente de haver ou não a indicação formal de falha humana, a indexação semântica latente detectou a presença da mesma através das frases presentes nos documentos D1, D2, D3, D4, D5 e D6, dentre outras.

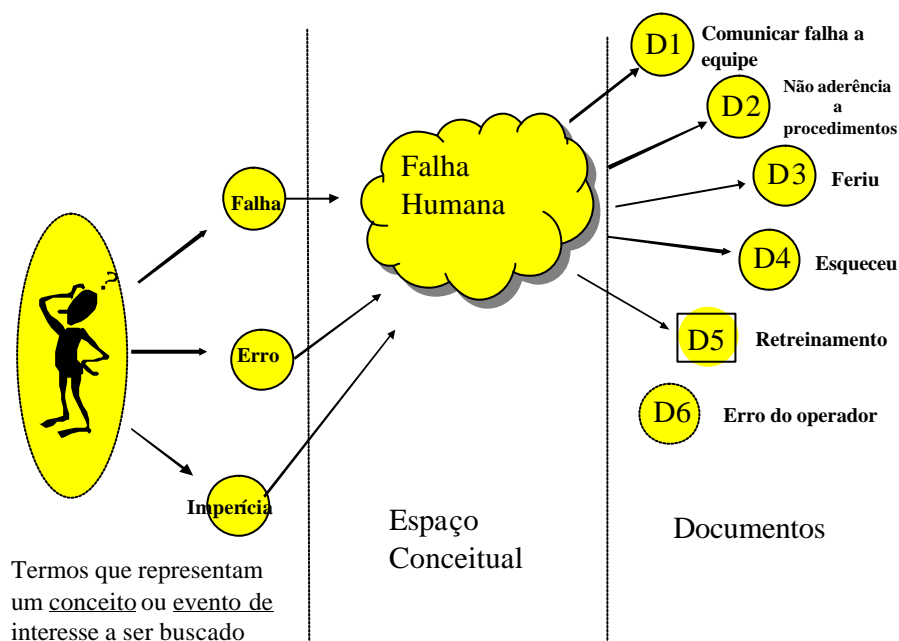



Figura 53: Aplicação de Indexação Semântica Latente em Mineração de Textos na Análise de Falha Humana na ELETRONUCLEAR

12 CONSIDERAÇÕES FINAIS

12.1 Avaliação da Trajetória da Pesquisa

Os documentos textuais são objetos portadores de informação e conhecimento facilmente elaborados por diferentes pessoas. A escrita é uma construção humana importante e hoje de acesso fácil à maioria das organizações em formato digital. Documentos digitais de texto são criados com facilidade atualmente, seja em editores de textos seja no ambiente Web. É um modo relativamente fácil e pouco formal para se converter parcelas de conhecimento tácito em conhecimento explicito. Obviamente a capacidade de expressão escrita possui suas limitações, como a da educação formal recebida por uma pessoa que lhe confere uma habilidade maior ou menor de escrever. Além disso, os pensamentos e reflexões do indivíduo circulam em sua mente a uma velocidade e intensidade bem maior do que a velocidade com que suas mãos digitam um texto em computador ou transcrevem palavras em um papel. Ainda assim, não é difícil para parcelas significativas das reflexões, insights, pensamentos e visões de uma pessoa saltem de sua mente para um texto, seja ele bem ou mal elaborado, grande ou pequeno, de caráter cientifico ou informal, estruturado ou em pequenos trechos de registro em formulários.

Por sua vez, conhecimento tácito de uma organização se constitui, por um lado, desafio gerencial e, por outro, em diferencial estratégico. Não se questionam o valor e a importância crescente desse conhecimento, especialmente nos ambientes que lidam com produtos e serviços de alto valor agregado ou dependentes do uso de tecnologia intensiva que demanda aplicação de avanços da ciência. Entretanto, sua captura e tratamento são complexos e onerosos para a organização, nunca ocorrendo em sua totalidade, uma vez que o conhecimento passível de ser expresso por um indivíduo é sempre inferior ao que ele efetivamente detém. Por outro lado, algumas ações são passíveis de implantação.

Em ambientes complexos, como os estudados na ELETRONUCLEAR, lidar com o conhecimento tácito – no caso em questão, a Experiência Operacional – se constitui em um desafio que pode ser respondido com a conciliação de estratégias relacionadas às pessoas, conforme descrito por RINTALA & KURONEN (2006) em

seus 23 métodos de compartilhamento de conhecimento tácito, bem como com a tecnologia de mineração de textos discutida nessa Tese. O resultado do emprego simultâneo dessas abordagens é um ação organizacional eficaz, sob o ponto de vista do emprego do conhecimento na ação empreendedora da organização, para que cumpra suas atividades-fim.

Não obstante o presente trabalho tenha sido realizado em uma organização específica, os aspectos estudados são comuns a muitas outras organizações com características semelhantes, ou seja: complexidade operacional, dependência de conhecimento e presença de muitas bases documentais digitais. A maior parte do aporte conceitual e tecnológico com recursos de inteligência artificial e métodos da área de recuperação da informação que estão embutidos no processo de descoberta de conhecimento em bases de textos são aplicáveis a coleções digitais de outras organizações. A variedade de métodos e técnicas disponíveis para serem embutidas em um software que realizará a mineração de dependerá também do domínio da aplicação e da população de usuários a ser atingida.

A descoberta de conhecimento em bases de textos, embora relevante enquanto estratégia e recurso tecnológico para a gestão do conhecimento, conforme constatado nessa tese, possui ainda muitas limitações. Em primeiro lugar, é preciso salientar que não se trata de tecnologia para venda em escala, ou seja: não permite gerar a oferta comercial de softwares de prateleira, facilmente disponibilizados e consumíveis por usuários. Ao contrário: seu emprego depende fortemente do domínio da aplicação, que merece ser estudado para que recursos metodológicos e computacionais possam ser alinhados visando o alcance da tarefa maior de descoberta de conhecimento.

Por outro lado, são realmente poderosos os recursos de certas tarefas de mineração de textos, especialmente a sumarização, categorização e *clusterização*, e atividades de refinamento das bases textuais, são potencialmente significativos para um grande número de organizações, desde que sintonizadas adequadamente com o domínio da aplicação, considerando as especificidades do conhecimento envolvido.

A ELETRONUCLEAR é uma organização que já avançou de modo significativo no processo de Gestão do Conhecimento, com ações concretas já tendo

sido implantadas, tais como a Gestão Eletrônica de Documentos e o Portal Corporativo de Experiência Operacional. São conquistas importantes no campo do uso de recursos para apoiar a Gestão do Conhecimento. A introdução da tecnologia de mineração de textos promoverá um salto qualitativo ainda maior para a organização. Seus profissionais, especialmente os detentores de experiência acumulada e alto nível de formação, com experiência operacional na condução de usinas nucleares poderão se beneficiar em larga escala dessa tecnologia, economizando parcelas de tempo em pesquisas e estudos sobre conhecimento nuclear conforme suas necessidades e realizando novas análises a partir de inferências permitidas pela análise sinóptica dos conjuntos de textos tratados pelo software ***DifBrain*®.**

A abordagem de mineração de textos é uma estratégia de tecnologia da informação, articulada com a recuperação da informação e uso de sistemas inteligentes que tem por objetivo ajudar e não substituir as pessoas na relação com o conhecimento. A própria expressão "descoberta de conhecimento em textos" poderia ser questionada, sob o ponto de vista de que se realmente é possível a uma ferramenta de software "descobrir conhecimento" seja em textos seja em dados. No entanto, a expressão é uma abstração, uma figura de linguagem para expressar a idéia de que um conjunto de recursos tecnológicos pode atuar de forma tão competente e precisa, articulando múltiplas tarefas de forma a simular uma habilidade que é exclusivamente humana que é lidar com o conhecimento, empregando-o conforme seus objetivos e seus valores. Uma ferramenta faz emergir certas percepções que seriam muito difíceis para um ser humano construir (como o faz a mineração de textos na categorização *clusterização*), mas conhecimento e sabedoria são características e virtudes eminentemente humanas.

Os pressupostos adotados no contexto dessa tese podem ser considerados confirmados, uma vez que:

- Os textos podem ser confirmados como reais expressões do pensamento humano, com grande flexibilidade e grau de liberdade para elaboração por parte das pessoas, tornando-se assim um modo acessível de conversão de conhecimento tácito para conhecimento explícito em muitos contextos organizacionais. Isso foi indicado na grande base de documentos possuída pela

ELETRONUCLEAR, em que há grande variação de estilos, formas, estruturas e grau de liberdade conferido ao usuário externo para elaboração dos textos. Por outro lado, a mineração de textos realmente confere às coleções textuais digitais armazenadas nas organizações um novo e significativo potencial de geração da informação e conhecimento, como foi demonstrado pelos experimentos realizados dos quais se constatou eclodir visões, percepções e consciência de novos conhecimentos não tão facilmente percebidos.

- Utilizou-se abordagens semânticas e estatísticas para responder de forma inovadora e eficaz as necessidades de geração de informação útil para o usuário. Nas coleções de textos digitais da ELETRONUCLEAR, as informações geradas pela ferramenta de mineração de textos, por sua capacidade de análise semântica e possibilidade de construção de visões sinópticas de grupos de documentos (derivadas especialmente dos processos de sumarização, categorização e *clustering*) se mostraram relevantes para a geração de conhecimento novo para o usuário, satisfazendo-o em suas necessidades. Por fim, o uso de recursos de inteligência artificial, revelou sua potencialidade e seus ganhos visíveis em ambientes complexos que incluam a necessidade de manipular grandes volumes de informação textual, uma vez que a ferramenta utilizada incorporou tais recursos.

Em síntese podemos afirmar que a presente Tese procurou entender, subsidiar e aprimorar o processo de gestão do conhecimento em organizações com o aporte da tecnologia de mineração de textos, e propôs a implantação de um modelo de recuperação automática da informação textual digital , ou descoberta de conhecimento em bases textuais digitais para a Experiência Operacional da ELETRONUCLEAR. Foram estudadas, refletidas/aperfeiçoadas e incorporadas, ao longo do tempo, as diversas contribuições teóricas e proposições de modelos e arquiteturas para sistemas informatizados de mineração de textos, com destaque para as de FELDMAN & SANGER (2006), KONCHADY (2006), ZANASI (2005), SULLIVAN (2001), REZENDE (2003) e WIVES (2002) que permitiram a definição de modelos conceituais, identificação e estudo de recursos de mineração de textos a serem implementados em uma ferramenta computacional e que foi construída e aprimorada ao longo de dois anos e meio de trabalho de desenvolvimento. Estudando as características das necessidades

de informação e geração e uso de conhecimento relativo à operação de usinas nucleoelétricas, concluiu-se pela pertinência e adequação da tecnologia de mineração de textos e do software desenvolvido que a embute: o ***DifBrain®*** .

12.2 Perspectivas Futuras

A presente Tese abre algumas perspectivas que podemos relacionar. Inicialmente, a indicação do valor e potencialidade da tecnologia de mineração de textos como suporte à gestão do conhecimento indica sua ampliação de aplicação. No campo da Experiência Operacional Interna, foi já identificado um outro foco de estudo que é sua utilização sobre a base de documentos concernente aos procedimentos operacionais e normas técnicas da ELETRONUCLEAR, que é extensa e demanda um trabalho constante de atualização, análise e disponibilização, exigindo avaliações de relacionamentos entre muitos documentos em geral articulados entre si, ou seja: a alteração em um documento implica em modificações em vários outros, por exemplo.

Um outro aspecto importante a ser considerado é ampliação de uso de modelos de Recuperação da Informação e dos Sistemas Inteligentes no processamento de textos (descritos nos capítulos 5 e 6 e que podem ampliar ainda mais a potencialidade da ferramenta de mineração de textos. A incorporação de um tesauro já foi considerada (existe um tesauro de termos para a área nuclear) para aprimorar o tratamento dos documentos.

Migrando da Experiência Operacional Interna para a Experiência Operacional Externa da ELETRONUCLEAR, cujas bases textuais tendem a crescer a taxas ainda maiores do que as bases internas de documentos, novos focos de análise inteligente podem ser desenvolvidos, englobando inteligência organizacional, monitoramento de tendências relativas a desempenho e segurança das usinas, acompanhamento de tendências internacionais, dentre outros aspectos. O uso de mineração de textos na Experiência Operacional Externa (que possui uma base de cerca de 20 mil documentos) pode contribuir para a análise de melhores práticas da organização. A tecnologia também pode ser estendida para às áreas administrativas da Empresa, tais como, marketing, recursos humanos, planejamento estratégico, dentre outras. Outras

organizações também podem se beneficiar profundamente da tecnologia de mineração de textos, conforme discutido por ZANASI (2005) e apresentado no capítulo 8 da presente Tese.

Outro campo potencial de crescimento da tecnologia estudada, é o que normalmente vem sendo denominado de *web mining*, ou seja: aplicação do processo de descoberta de conhecimento em bases de documentos digitais presentes na *Web*, envolvendo a análise de conteúdo, da estrutura e do uso da *Web* (REZENDE, 2003). *Web mining* de conteúdo é o processo de descoberta de conhecimento relativo ao conteúdo das páginas e de suas descrições. *Web mining* de estrutura é o esforço por descobrir conhecimento a partir da organização da *Web* e da referência cruzada de ligações. Por fim, *Web mining* de uso é o processo de geração de padrões interessantes com o uso dos registros de acesso da *Web*.

Nos próximos anos, devem se intensificar os esforços pelo uso dessa tecnologia, que deverá ser alavancada ainda mais pelas contribuições do desenvolvimento e pesquisa na área de processamento da linguagem natural – PLN. Partindo da idéia de que o conhecimento a ser utilizado em um sistema inteligente, (como é o caso da mineração de textos), se encontra disponível em modo texto, a questão está em capturar e mapear tal conhecimento sob algum tipo de formalismo para representação neste sistema. A extração de informação de textos é uma questão abordada em grande parte na área de PLN. Extração de taxonomias semânticas, estudos sobre relações léxicas para representações de texto, categorização de mensagens de correio eletrônico, tratamento de ambigüidade léxico-morfológica, interpretação de estruturas implícitas no texto estão entre as contribuições dessa área para a tecnologia de mineração de textos (MONTOYO, MUÑOZ & MÉTAIS, 2005). O esforço para um tratamento semântico de bases textuais é hoje tônica de muitas pesquisas e irá contribuir muito para novos aperfeiçoamentos e uma nova geração de ferramentas para mineração de textos. Portanto, as pesquisas relacionando *text mining* com processamento de linguagem natural estão na essência de um processo de recuperação “inteligente” de informações e descoberta de conhecimento em textos, devendo crescer em relevância nos próximos anos.

REFERÊNCIAS

AAKER, D. _Marketing research_. New York: Wiley, 1990.

ALLEN, J. _Natural Language Understanding_. The Benjamin/Cummings Publishing Company, Inc., 1994.

ABITEBOUL, S., P. BUNEMAN, et alli. _Data on the Web: from relations to Semistructured Data and XML_. Delphi. San Francisco, 2000

ANDERSON, J. D., PÉREZ-CARBALLO, J. _The nature of indexing: how humans and machines anayze messages and texts for retrieval_. _Information Processing & Management,_ 37, 2001, 231-237

ARAÚJO, V.M.R.H. Sistemas de Recuperação da Informação: nova abordagem teórico conceitual. Tese de Doutorado. CFCH. Universidade Federal do Rio de Janeiro. Rio de Janeiro, 1994

BAEZA-YATES, R.; RIBEIRO-NETO, B. _Modern Information Retrieval_. New York: ACM Press, 1999.

BARBIERI, Carlos _Business Intelligence_: modelagem e tecnologia. Rio de Janeiro. Axcel Books, 2001

BELKIN E ROBERTSON _Information science and the phenomenon of information._ JASIS, v. 27, p.197-204, 1976

BERTALANFFY, Ludvig von. _General Systems Theory_. Nova York, George Brasilier, 1968

BRIET, Suzanne. Qu`est-ce que la documentation. Paris [s.n.] 1951

BRAGA, G. M. & CHRISTOVÃO, H. T. Ciência da Informação e Sociologia do Conhecimento Científico: a intertematicidade plural (sobre “A ciência e seu público”, de Lea Velho: um ponto de vista da Ciência da Informação). _Transinformação_, Campinas, v. 9, n. 3, p. 33-45, set./dez. 1997.

BOOCH, G., RUMBAUGH, J & JACOBSON, I. UML: Guia do Usuário. Rio de Janeiro. Campus, 2000
BORKO,H. _Information science: what is it?_ American Documentation v.19,n.1,p.3-5,1968.

BROOKSHEAR, J.G. Ciência da Computação: uma visão abrangente. Porto Alegre, Bookman, 2000

BRYMAN, A. _Research methods and organization studies_. Unwin Hyman. London, 1989

BUCKLAND, Michael _Information as a thing._ Journal of the American Society for Information Science, v. 42, n.5, p. 351-360, 1991

BUCKLAND, Michael *What is this a "document"?* Journal of the American Society for Information Science, New York: John Wiley & Sons. v.48, n.9, p.804-809, 1997

BURKE, P. Uma história social do conhecimento: de Gutemberg a Diderot. Rio de Janeiro, Jorge Zahar Editor, 2003

BURNETT, Kathleen; McKINLEY, E. Graham. *Modelling information seeking. Interacting with computers*, London, n. 10, p. 285-302, 1998.

BURKOWSKI, F. *Retrieval activities in a database consisting of heterogeneous collections of structured text. In: Proc of the 15th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, p 112-125. Copenhagen, Denmark, 1992.

BUSH, Vannevar *As We May Think* Atlantic Monthly, 1945

CALDER, B. *Focus group and the nature of qualitative marketing research*. *Journal of Marketing Research*, n. 14, p. 353-64, Aug. 1977.

CALLAN, James P. *Passage-level evidence in document retrieval*. In: CROFT, W. Bruce; RIJSBERGEN, C. J. van (eds). *VII International ACM-SIGIR Conference on Research and Development in Information Retrieval*. London: Springer-Verlag. 1994.

CANÇADO, Márcia. Manual de Semântica: noções básicas e exercícios. Belo Horizonte. Editora da UFMG, 2005.

CAPLAN, S. *Using focus group methodology for ergonomic design*. *Ergonomics*, v. 33, n. 5, p. 527-533, 1990.

CAVALCANTE, M., GOMES, E. Caderno de Projeto do curso de pós-graduação latu senso em Gestão do Conhecimento e Inteligência Empresarial(MBKM). Rio de Janeiro. Centro de Referência em Inteligência Empresarial da COPPE/UFRJ, 2002.

CELLÉRIER, G. *El pensamiento de Piaget* Barcelona, Península, 1976

CENDON, Beatriz V. Sistemas e Redes de Informação *in* OLIVEIRA, Marlene, et alli. Ciência da Informação e Biblioteconomia : novos conteúdos e espaços de atuação. Belo Horizonte. Editora UFMG, 2005

CHIAVENATO Idalberto. Introdução à Teoria Geral da Administração. 5ª. ed. São Paulo, Makron Books, 1976

CHOO, Chun W. A Organização do Conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: SENAC Editora, 2003

CHOUDHURY, V. e SAMPLER, J. L. *Information specificity and enviromental scanning: an economic perspective*. *MIS Quartely*. Março de 1997

CLAYBROOK, Billy G. Técnicas de gerenciamento de arquivos. 2ª ed. Rio de Janeiro. Campus, 1987

CNEN (2001) - Comissão Nacional de Energia Nuclear. Relatórios de Operação e Usinas Nucleoelétricas. CNEN-NN-1.14. Resolução N°016 de 29/11/2001. Disponível em http://www.cnen.gov.br/seguranca/normas/nn_114.doc. Acesso em 02 fev. 2007. Publicado no D.OU. em 10/01/2002.

CODD, Edgar F. *A Relational Model of Data for Large Shared Data Banks. Communications of the ACM – Association for Computing Machinery,* New York, v.13, n.6, p.377-387, jun. 1970.

COLANGELO FILHO, Lucio. Implantação de Sistemas ERP: um enfoque de longo prazo. São Paulo. Atlas, 2001

COVE, J.F., WALSH, B.C. *On line text retrieval via browsing*. Information *Processing & Management*, v.24, n. 1, p 31-37, 1988

DAVENPORT, T, *Knowledge Management Case Study – Knowledge Management at Microsoft*, Boston. Mc Combs School of Business, 1997

DAVENPORT, T. & PRUSAK, L Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual. Tradução de Lenke Peres. Rio de Janeiro: Campus, 1998

DIAS, Cláudia A. Portal corporativo: conceitos e características. Ci. Inf., Brasília, v. 30, n. 1, 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100007&lng=pt&nrm=iso>. Acesso em: 20 Dez 2006.

DILTHEY, W. Das Wesen der Philosophie, in: Kultur der Gerenwart I, VI: 3ª. ed. 1921

DORIGO M. & GAMBARDELLA, L. M. *Ant Colonies for the Traveling Salesman Problem. BioSystems, 43:73-81. Also Tecnical Report TR/IRIDIA/1996-3, IRIDIA*, Université Libre de Bruxelles, 1997

DORIGO, M.; STÜTZLE, T. *Ant Colony Optimization* The MIT Press, 2006

DRUCKER, Peter F. *The coming of the new organization*. Boston, Harvard Business Review. V.28, n.1, p.45-53, 1988

ELETRONUCLEAR. Informações gerais sobre a empresa e a energia nuclear no endereço eletrônico: www.eletronuclear.gov.br. Acessos diversos entre fevereiro de 2004 e fevereiro de 2007.

FELDMAN, R.; DAGAN, I. *"Knowledge Discovery in Textual Databases (KDT)". In First International Conference on KnowledgeDiscovery (KDD´95).* Montreal, l995.

FELDMAN, R., SANGER, J. *The Text Mining Handbook: Advanced Approachs in Analyzing Unstructured Data*. New York. Cambridge University Press, 2006

FLEURY, A . & FLEURY, M. Aprendizagem e Inovação Organizacional: as experiências de Japão, Coréia e Brasil São Paulo: Atlas, 1997

FORONDA, Diego A. H. Estudo exploratório da indexação semântica latente e das funções "peso". Dissertação de Mestrado. Porto Alegre. Faculdade de Informática. PUC-RS, 2005.

FOSKETT, A. C. *The Subject Approach to Information*. 5. ed. Londres: Library Association Publishing, 1997.

FRAKES, William B. *Introduction to information storage and retrieval systems*. In FRAKES, William B.; BAEZA-Yates, Ricardo A. *Information Retrieval: Data Structures & Algorithms*. Upper Saddle River, New Jersey: Prentice Hall PTR, 1992. p 1-12.

FRANÇA, Fabrício *Ant Colony Optmization e Swarm Intelligence*. Apresentação. Faculdade de Engenharia Elétrica e Computação. UNICAMP. 2006. Disponível em: www.dca.fee.unicamp.br/~olivetti. Acessado em: 20 jan. 2007

GARCIA MARCO, F.J.. *On some contributions of the cognitive sciences and epistemology to a theory of classification*. Knowledge Organization, v.20, n.3, 1993.

GIUSTA, Agnela. Concepções de Aprendizagem e Práticas Pedagógicas Revista de Educação, Belo Horizonte (1985: 24-31)

GOEBEL, Michael e GRUENWALD, Le. *A survey of data mining and knowledge discovery software tools*. *ACM SIGKDD Explorations*, v.1, n. 1, 1999. Disponível em: http://www.acm.org/sigs/sigkdd/explorations/issues/1-1-1999-06/survey.pdf. Acessado em 12/09/2006.

GORDON, M.D., DUMAIS, S. *Using latent semantic indexing for literature based discovery*. *Journal of the American Society for Information Science*, 49, 1998, 674-685.

HABERMAS, G. Conhecimento e interesse. Rio de Janeiro: Zahar, 1982.

HESSEN, J. Teoria do Conhecimento . Trad. João Virgílio Cuter. São Paulo. Martins Fontes 1999

HUSSERL, E. *Logische Untersuchungen*. Halle. Niemeyer, 1913

IAEA – *International Atomic Energy Agency*. *Senior Level Meeting on Managing Nuclear Knowledge – Introductory Discussion*. Paper, 2002

IANNI, O. A sociedade global. Rio de Janeiro: Civilização Brasileira, 1992.

INMON, William H. Buiding the data warehouse. John Wiley & Sons, Inc. USA, 1996

JARVENPAA, Sirkka L. TANRIVERDI, *Huseyin. Leadin virtual konwledge networks*. Creativity and Innovation Management, v.8, n.2 p. 70-76, Jun. 1999

JOHANNESSEN, J.A., OLAISSEN, J., & OLSEN, B. *Mismanagement of tacit knowledge: the importance of tacit knowledge, the danger of information technology an what to do about it*. *International Journal of Information Management*. Vol 28. p. 3-20.

KIMBALL, Ralph. *The Data Warehouse Toolkit*. John Wiley & Sons, Inc. USA, 1996

KIMBALL, Ralph. *The Data Warehouse Lifecycle Toolkit*. New York J. Wiley & Sons, 1998

KONCHADY, M. *Text Mining application programming* Boston, MA, USA. Thomson Learning Inc., 2006

KORTH, H. SILBERSCHATZ, A. & SUDARSHAN, S. Sistema de Banco de Dados. 3ª. Edição Brasileira. São Paulo. Makron Books, 1999.

KUHLTHAU, Carol *Inside the search process: information seeking from the user's perspective*. *Journal of the American Society for Information Science*, v. 42, n. 5, junho de 1991.

KURAMOTO, Hélio. Sintagmas Nominais: uma Nova Proposta para a Recuperação de Informação**.** em *DataGramaZero - Revista de Ciência da Informação - v.3 n.1 fevereiro* 2002

KURZ, R. O colapso da modernização: da derrocada do socialismo de caserna à crise da economia mundial. Rio de Janeiro. Paz e Terra, 1993.

KURZ, R. A ignorância na sociedade do conhecimento. Folha de São Paulo. Caderno Mais. 20 de maio de 2002.

LANCASTER, F. Wilfrid. *Information Retrieval Systems: Characteristics, Testing and Evaluation*. New York: John Wiley & Sons, 1968.

LANCASTER, F. W. O currículo da Ciência da Informação. *Revista de iblioteconomia de Brasíli*a, Brasília, v. 17, n.1, p. 01-05, jan./jun. 1989.

LANCASTER, F. Wilfrid Indexação e Resumo: teoria e prática. 2ª. Edição Brasileira. Briquet de Lemos, Brasília, 2004.

LANCASTER, F. W. e WARNER, A. J. *Information Retrieval Today*. Information Resources Press, 1993.

LARMAN, Craig. Utilizando UML e Padrões: uma introdução à análise e ao projeto orientados a objetos. Porto Alegre. Bookman, 2000.

LE COADIC, Yves-François A Ciência da Informação. Brasília. Briquet de Lemos, 2004.

LE GUERN, Michel. *Traitement automatique et variation linguistique : la syntaxe des titres. In: OPÉRATEURS et Constructions Syntaxiques : évolutions des marques et des distributions du Xvème siècle. Paris : Presses de l'Ecole Normale Supérieure*, 1994. p. 75-81.

LEPECKI, W. P. Determinação do Know-How Tecnológico da Eletronuclear.. Angra dos Reis – RJ. ELETRONUCLEAR, 2002.

LIEBSCHER, P. *Quantity with quality ? Teaching quantitative and qualitative methods in a LIS Master's program. Library Trend*s, v. 46, n. 4, p. 668-680, Spring, 1998.

LINTHICUM, David S. *Enterprise Application Integration. Information technology series*. Addison Wesley, 2000.

LOH, Stanley. Uma abordagem para busca contextual de documentos na Internet. Porto Alegre. Revista de Informática Teórica e Aplicada (RITA), v.4, n.2, p. 79-92, 1997

MACHADO, A. M. N. *Informação e controle bibliográfic*o: um olhar sobre a cibernética. Tese (Doutorado em Educação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2001

MARTINS, J.C. Gerenciando projetos de desenvolvimento de software com PMI, RUP e UML. Rio de Janeiro: Brasport, 2004

MARX, K. O capital, Livro I. Trad. port., Rio de Janeiro: Civilização Brasileira, 1968.

MCDANIEL, Carl & GATES, Roger. Pesquisa de Marketing. São Paulo. Pioneira-Thomson, 2003.

MEADOWS, Charles *Text Information retrieval systems*. San Diego. Academic Press, 1992

MERLEAU-PONTY, M. *Phénoménologie de la perception.* Gallimar. Paris, 1945.

MESSIAS, L.C.S. & MORAES, J.B.M. Informação: conceitos e terminologias na área de Ciência da Informação. ENANCIB, Anais. Belo Horizonte, 2003

MEYER, M., BABER, R. e PFAFFENBERGER, B. Nosso Futuro e o Computador Porto Alegre, Bookman: 2000

MIZZARO, Stefano. *A Cognitive Analysis of Information Retrieval*. In: *Information Science: integration in perspective* - CoLIS2, 1996, Copenhagem, Denmark. *Proceedings... The Royal School of Librarianship*, 1996. p.233-250. Disponível em: http://ten.dimi.uniud.it/~mizzaro/papers/colis.pdf. Acesso em: 01.set.2006

MIZZARO, Stefano. Relevance: *The Whole History*. *Journal of the American Society for Information Science*, New York, v.48, n.9, p. 810-832, 1997.

MIZZARO, Stefano. *How many relevances in information retrieval? Interacting with Computers*, London, n. 10, p. 303-320, 1998.

MONTGOMERY, Douglas C. *Design and Analysis of Experiments. 5 th Ed.:* New York. John Wiley, 1997.

MONEY, Raymond J. *Course CS 371R: Information Retrieval and Web Search - Slides*. Austin, TX, USA. University of Texas at Austin. 2006. Disponível em http://www.cs.utexas.edu/~ mooney. Acesso em 01. fev. 2007

MONTOYO, A. , MUÑOZ, R. & MÉTAIS, E. et alli. *Natural Language Processing and Information Systems. 10th International Conference on Applications of Natural Language to Information Systems, NLDB. Alicante, Spain, June 2005, Proceedings*. Berlin, Germany. Springer-Verlag, 2005.

MOORES, Calvin N. *Datacoding applied to Mechanical Organization of Knowledge*. American Documentation, USA . v.2, p.20-32. 1951.

MORAES NETO, B.R. A organização do trabalho sob o capitalismo e a redoma de vidro. Revista de Administração de Empresas da FGV. 27 (4) 19-30 Rio de Janeiro, FGV, 1987.

MORRIS, J; Hirst, Graeme. *Lexical Cohesion Computed By Thesaural Relations As An Indicator Of Th Structure Of Text*. Computationallinguistics, V.17, N.1, Março de 1991.

MORSE, P.M. Browsing and search theory. In: RAWSKI, C.H. Toward a theory of librarianship; paper in honor of Jesse Hank Shera. Melachen, N.J.: The Scarecrow Press, 1973. 564 p., p.246-261.

NADLER, David et alli. Arquitetura Organizacional: a chave para a mudança empresarial. Rio de Janeiro. Editora Campus, 1994

NEHMY, R.Q. O ideal do conhecimento codificado na 'era da informação: o programa de gestão do conhecimento. Tese de Doutorado. Belo Horizonte. Escola de Ciência da Informação da Universidade Federal de Minas Gerais, 2001.

NETO, Roberto. *Mapeamento entre os modelos E/R e Star Paper.* Rio de Janeiro, DCC/NCE/UFRJ 1998.

NEVES, Jorge T.R e NASCIMENTO, Niraldo J.N. A Gestão do Conhecimento na Word Wide Web: reflexões sobre a pesquisa de informações na rede. In. Perspectivas em Ciência da Informação, v.4, n. 1, p. 29-48, jan./jun.1999.

NONAKA, I. e TAKEUCHI, H. Criação do Conhecimento na Empresa: como as empresas japonesas geram a dinâmica da inovação. Tradução de Ana Beatriz Rodrigues e Priscilla Martins Celeste. Rio de Janeiro. Editora Campus, 1997.

NOVINS, P., ARMSTRONG, R., *Choosing Your Spots for Knowledge Management*. New York. *Ernst & Young Center of Business Innovation, Innovation Journal, Issue* 1, v. 8, pp 45-52, 2002. Disponível em: http://www.providersedge.com/docs/km_articles/Choosing_Your_Spots_for_KM.pdf. Acessado em 31.10.2006

O'BRIEN, A. J. Sistemas de Informação e as Decisões Gerenciais na Era da Internet. São Paulo, Saraiva, 2001

OLIVEIRA (1), Marlene, Origens e Evolução da Ciência da Informação. *in* OLIVEIRA, Marlene *et alli.* Ciência da Informação e Biblioteconomia : novos conteúdos e espaços de atuação. Belo Horizonte. Editora UFMG, 2005

OLIVEIRA (2), Luiz C. Gestão do Conhecimento na Eletronuclear. Projeto Final. Especialização. Rio de Janeiro. COPPE-Universidade Federal do Rio de Janeiro, 2002

OLIVEIRA (2), Luiz C. Portal de Experiencia Operacional e Informação Tecnológica da Eletronuclear. Slides. Angra dos Reis. Eletronuclear, 2006

OLIVEIRA, L, FILHO, D e SOUZA, M. Gerenciamento Eletrônico de Documentos: Caso prático da Eletronuclear. Belo Horizonte. V SINCONEE – Simpósio de Gestão do Conhecimento do Setor de Energia Elétrica. Apresentação. 2004

PAO, M.L. *Concepts of information retrieval*. Englewood, Cols.: Libraries Unlimited. 1985, p. 65

PIAGET, Jean. *Le structuralism*. Paris. Presses Universitaires de France, 1968

PIAGET, Jean. A equilibração das estruturas cognitivas. São Paulo. Zahar, 1976

PIRES DE OLIVEIRA, R. Semântica Formal: uma breve introdução. Campinas. Mercado de Letras, 2001

POLANYI, M *Knowing and Being.* Chicago University Press, 1969

POLANYI, M. *Personal Knowledge:* towards a post-critical philosophy. Chicago, The University of Chicago Press, 1962.

POLANYI, M. The tacit dimension. London: Routledge & Kegan Paul, 1966.

PORTER, Michael. Vantagem Competitiva: criando e sustentando um desempenho superior. Rio de Janeiro: Campus, 2000

PRADO, H. A. Conceitos de Descoberta de Conhecimento em Banco de Dados. Trabalho Individual. Porto Alegre. CPGCC – UFRGS, 1997

REES, A. & SARACEVIC, T. *Conceptual analysis of questions in information retrieval systems.* Procedings of the Annual meeting of the American Documentation Institute, 1963, p. 175-177

REZENDE, S. et alli. Sistemas Inteligentes: fundamentos e aplicações. São Paulo, Manole, 2003

RIBEIRO, J.L.D.; FOGLIATO, Flávio S. and CATEN, Carla S. T. Minimizing Manufacturing and Quality Costs in Multiresponse Optimization. Porto Alegre. *Quality Engineering*, 13(2), 2000.

RIJSBERGEN, C. *Information Retrieval*. 2ed. London: Butterworths, 1979.

RINTALA, Niina e KURONEN, Tanja *How to share tacit nuclear knowledge? Paper. In. J. Nuclear Knowledge Management*. Vol. 2. No. 2, 2006

ROBREDO, Jaime e CUNHA, Murilo. *Documentação de Hoje e de Amanhã: uma abordagem informatizada da biblioteconomia e dos sistemas de informação*. 2. ed. São Paulo. Global, 1994.

RUSSEL, Stuart. NORVIG, Peter. Inteligência Artificial. 2a. Ed. Rio de Janeiro. Campus, 2004.

SALTON, Gerard; MACGILL, Michael J. *Introduction to Modern Information Retrieval*. New York: McGRAW-Hill, 1983.

SALTON, Gerard. *A note about Information Science research*. in BOYCE, Bert, KRAFT, Donald H. Principles and theories in Information Science. Annual Review of Information Science and Technology *(Arist*). v.20, p.153-178, 1985.. p.155.

SARACEVIC, Tefko. *The concept of "relevance" in Information Science: an historical review*. In:. *Introduction to Information Science*. New York, R. R. Bowker Co., 1970. p. 111-154.

SARACEVIC, Tefko. On a method for studying the structure and nature of requests in information retrieval. Proceedings of the ASIS, v. 20, p. 22-25, 1983

SARACEVIC, Tefko. *Information science: origen, evolution and relations*.In: VAKKARI, P., CRONIN, B. eds. Conceptions of Library and Information Science. Proceedings of the COLIS Conference Tampère, 1991. Los Angeles, Taylor Grahan,1992.

SCHUMPETER, J.. A teoria do desenvolvimento econômico. Trad. port., São Paulo: Nova Cultural, 1985.

SCHEIN, E.H., Process Consultation, vol 1,. Addisson-Wesley Publishing Company, Inc, Cap. 6. 1986

SERRA, Laércio A essência do *business intelligence*. São Paulo, Berkeley, 2002

SETZER, V. W. Dado, informação, conhecimento e competência. *Data Grama Zero – Revista de Ciência da Informação*, Rio de Janeiro, n. zero. dez. 1999. Disponível em: http://www.ime.usp.br/~vwsetzer/datagrama.html Acesso em: 02 jul. 2006.

SHANNON, C. E. *A Mathematical Theory of Communication*. [S.l.]: Bell Labs, 1998. Disponível em: < http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html >.Acesso em: 01 fev. 2005.

SHNEIDERMAN, Ben. *Universal usability. Communications of the ACM*, v. 43, n. 5, p. 85-91, 2000

SILVA, E.C. Sistemas de gerenciamento de documentos para centro de documentação e informação. Dissertação de Mestrado. Belo Horizonte: Escola de Biblioteconomia da UFMG, 1995.

SIQUEIRA, André. Potência Nuclear? In: Revista Carta Capital, São Paulo (410) 10-13, setembro, 2006.

SMITH, A. A riqueza das nações: investigação sobre sua natureza e suas causas. Trad. port., São Paulo: Abril Cultural, 1983.

SOUTO, Kelling C. Sistema especialista com lógica nebulosa para o cálculo em tempo real de indicadores de desempenho e segurança na monitoração de usinas nucleares. Tese de Doutorado. Rio de Janeiro. COPPE/UFRJ, Engenharia Nuclear, 2005

SOUZA, Renato. Sistemas de recuperação de informações e mecanismos de busca na web : panorama atual e tendências. *in* Perspect. ciênc. inf., Belo Horizonte, v.11 n.2, p. 161 -173, mai./ago., 2006.

STEWART, Thomas. Capital intelectual - a nova vantagem competitiva das empresas. Rio de Janeiro: Campus, 1998.

SULLIVAN, Dan *Document Warehousing and Text Mining: techiniques for improving business operations, marketing and sales*. New York, Wiley, 2001

TERRA, José C. C. Gestão do conhecimento - o grande desafio empresarial - uma abordagem baseada no aprendizado e na criatividade. São Paulo: Negócio, 2000.

THIOLLENT, Michel. Metodologia da Pesquisa - ação. 3a ed. São Paulo: Cortez Editora, 1986.

THOMSEN, Erik. *OLAP Solutions: building muldimensional information systems.* New York. Jonh Wiley & Sons, Inc., 1997

TURBAN, Efraim Administração de Tecnologia da Informação. Rio de Janeiro, Campus, 2003

VAN MAANEN, J. *Reclaming qualitative methods for organizational research: a preface*. Administrative Science Quartely, v.24, n.4, Dec 1979, p.520-537

VAUGHN, S. et al. *Focus group interviews in education and psychology*. Thousand Oaks, CA: Sage Publications, 1996.

VICKERY, B & VICKERY, A. Information Science in theory and practice. Londres. Butterworths, 1987

WEIXUAN, X., ZHENGXIN, C., YONG, S. Org. *Data Mining and Knowledge Management: Chinese Academy of Sciences Symposium on Data Mining and Knowledge Discovery CASDMKD 2004, held in Beijing, China*. New York. Springer Verlag, 2004

WESTPHAL, C. BLAXTON, T. *Data Mining Solutions: methods and tools for solving real-word problems*. N.Y. Jonh Wiley & Sons, Inc., 1998

WIVES, L.K. Tecnologias de Descoberta de Conhecimento em Textos Aplicadas à Inteligência Competitiva. Exame de Qualificação. Porto Alegre. PPGC-Universidade Federal do Rio Grande do Sul, 2002

ZADEH, Laft A. *Fuzzy Sets*. *Information and Control*, v.8, n.1, p.338-353. 1965.

ZANASI, Alessandro et alli. *Discovering data Mining*. New York: Prentice Hall, 1997

ZANASI, A. et alli. *Text Mining and its Applications to Intelligence, CRM and Knowledge Management*. Southampton, United Kingdom, WIT Press, 2005

ZEMAN, J. Significado filosófico da noção de informação. In: ROYAUMONT, C. de (Org.). O conceito de informação na ciência contemporânea. Rio de Janeiro: Paz e Terra, 1970. p. 154–168.

ANEXO I

RESUMO DO RELATÓRIO PROJETO DETERMINAÇÃO DO .“KNOW-HOW” TECNOLÓGICO DA ELETRONUCLEAR [106]

1 – SISTEMÁTICA DE CLASSIFICAÇÃO DO KNOW-HOW

O projeto teve início com a busca de uma forma de classificar o “know-how” da empresa. Existem várias sistemáticas de classificação de “know-how” na ELETRONUCLEAR, mas cada uma com uma certa finalidade específica. Em particular, elas são fortemente vinculadas à usina considerada (Angra 1 ou Angra 2). Assim, decidiu-se por criar uma sistemática nova, adaptada à realidade da empresa. Para tanto, foram consultados Diretores, Superintendentes, Gerentes e especialistas, sendo o produto obtido fruto de uma atividade interativa ocorrida ao longo de sete meses (março a outubro de 2001). A Sistemática de Classificação de Know-How Tecnológico da ELETRONUCLEAR, assim obtida, está dividida em quatro grandes conjuntos:

1) Engenharia de Projeto e Apoio à Operação,
2) Realização Física do Empreendimento,
3) Operação
4) Atividades de Sustentação da Empresa.

Estes grandes temas foram sendo subdivididos em níveis sucessivos de detalhamento. O número de níveis foi deixado a critério de cada área, não havendo uniformidade quanto a este aspecto, e tendo-se chegado a um total de 557 itens. A seguir apresentamos um exemplo (Projeto de Sistemas Mecânicos) retirado da sistemática de classificação em questão:

[106] LEPECKI, W. P. Determinação do Know-How Tecnológico da Eletronuclear.. Angra dos Reis – RJ. ELETRONUCLEAR, 2002. Esse anexo foi apresentado como Apêndice I em: OLIVEIRA (2), Luiz C. Gestão do Conhecimento na Eletronuclear. Projeto Final. Especialização. Rio de Janeiro. COPPE-Universidade Federal do Rio de Janeiro, 2002

1.1 PROJETO DE SISTEMAS MECÂNICOS

1.1.1 Projeto dos Sistemas do Primário

1.1.1.1 Circuito Primário

1.1.1.2 Sistema de Controle Químico e Volumétrico

1.1.1.3 Sistemas de Remoção de Calor Residual

1.1.2 Projeto dos Sistemas Nucleares Auxiliares

1.1.2.1 Sistemas de Tratamento e Armazenamento do Refrigerante Primário

1.1.2.2 Sistemas de Refrigeração de Componentes

1.1.3 Projeto de Sistemas do Secundário

1.1.3.1 Circuito Secundário

1.1.3.2 Sistema de Água de Alimentação de Emergência

1.1.3.3 Sistemas de Água de Refrigeração

1.1.3.4 Sistemas Auxiliares Convencionais

1.1.4 Projeto de Sistemas Químicos

1.1.4.1 Sistemas de Tratamento de Água

1.1.4.2 Sistemas de Amostragem e de Dosagem de Produtos Químicos

1.1.5 Sistemas de Tratamento de Rejeitos

1.1.5.1 Sistema de Tratamento de Rejeitos da Usina

1.1.5.2 Depósitos de Rejeitos no Sítio

2 – DESCRIÇÃO SUMÁRIA DA METODOLOGIA DE LEVANTAMENTO DO “KNOW-HOW” TECNOLÓGICO DA ELETRONUCLEAR

A metodologia adotada para o levantamento foi uma auto-avaliação através da cadeia hierárquica da empresa, abrangendo todos os níveis gerenciais. Para execução desta auto-avaliação utilizou-se um questionário respondido pela *Intranet* da empresa através do formulário eletrônico desenhado para tal fim. A partir do preenchimento do formulário, foi feita uma avaliação quanto ao conhecimento explícito documentado, identificado no relatório como *Informação Tecnológica*, e quanto ao conhecimento tácito, identificado no relatório como *Pessoal Capacitado.* A situação de cada atividade, hoje e prospectivamente em cinco anos, foi avaliada com relação a disponibilidade e atualização da informação documentada, através da atribuição de um grau obtido pela média ponderada dos graus atribuídos a cada tipo de documento pelo gerente diretamente responsável pela atividade, considerando um peso que leva em conta a relevância do tipo de documento para a execução da tarefa. Os critérios para aplicação dos graus e dos pesos foram os seguintes:

Graus:

NA - Não aplicável
0 - Inexistente
1 - Existente parcialmente e desatualizada
2 - Existente totalmente e desatualizada
3 - Existente totalmente e atualizada
4 - Existente totalmente e acima das necessidades

Pesos:

0 - Desnecessária / Irrelevante
1 - Pouco necessária / Pouco relevante
2 - Necessária / Relevante
3 - Muito necessária / Muito relevante

Cada atividade também foi avaliada hoje e prospectivamente em cinco anos, quanto a disponibilidade de pessoal com a qualificação necessária e na quantidade requerida, através da atribuição de um grau, obtido pela média aritmética dos dois graus atribuídos

pelo gerente diretamente responsável pela atividade, seguindo os critérios descritos abaixo.

Qualitativamente - Graus:

NA - Não aplicável

0 - Não há nenhuma capacitação
1 - A capacitação do grupo é fraca
2 - A capacitação do grupo é mediana
3 - A capacitação do grupo é alta / elevada
4 - A capacitação do grupo está acima das necessidades

Quantitativamente – Graus:

0 - Não há ninguém trabalhando
1 - O número de pessoas é insuficiente / estamos deixando de fazer trabalhos /estamos fazendo muitas horas extras
2 - O número de pessoas é razoável / estamos fazendo os trabalhos no nosso limite / às vezes temos que fazer horas extras
3 - O número de pessoas é suficiente / temos folga para fazer as atividades e nos atualizar
4 - O número de pessoas ultrapassa as necessidades da atividade

Cada formulário corresponde a um item do Know-How Tecnológico, conforme a classificação, já mencionada, criada especialmente para o projeto. Os formulários foram preenchidos no nível mais baixo de subdivisão, pelas áreas (gerências) responsáveis, resultando em um total de 557 formulários. Os questionários foram distribuídos a 41 gerentes, havendo um índice de retorno de 92%. A informatização do processo permitiu uma resposta rápida (2 meses).

Os resultados estão armazenados em um banco de dados, o que permite fazer vários tipos de relatórios, de acordo com diferentes critérios, a partir dos quais análises e propostas de ações serão executados. Estes relatórios são criados no computador,

diretamente na tela dos usuários (diferentes níveis de gerência). De posse dos resultados a alta administração da empresa priorizou as atividades segundo a sua relevância para o negócio atribuindo um grau de 1 a 3, sendo 3 a mais alta prioridade, e 1 a mais baixa. Após o levantamento, e a priorização, foram identificadas as áreas em situação crítica, ou seja aquelas que satisfaziam um dos dois seguintes critérios:

a) Informação Tecnológica

- Grau máximo : 1-Existente parcialmente e desatualizado
- Peso mínimo: 3- Muito necessário/muito relevante

b) Pessoal Capacitado

- Grau máximo: 1- Capacitação fraca/ Pessoal insuficiente.

As situações consideradas críticas se transformaram em alvo de um estudo mais detalhado para elaboração de um plano de ação que assegure a retenção dos conhecimentos essenciais para execução das tarefas.

ANEXO II

Norma CNEN

Relatórios de Operação e Usinas Nucleoelétricas

Resolução Nº 016 29/11/2001

Publicação D.O U. 10/01/2002

SUMÁRIO

CNEN-NN-1.14 RELATÓRIOS DE OPERAÇÃO DE USINAS NUCLEOELÉTRICAS

Disponível em: http://www.cnen.gov.br/seguranca/normas/nn_114.doc.

RELATÓRIOS DE OPERAÇÃO DE USINAS NUCLEOELÉTRICAS

1 OBJETIVO E CAMPO DE APLICAÇÃO

1.1 OBJETIVO

Esta Norma tem por objetivo estabelecer os requisitos do programa de notificações de *eventos* significativos e dos relatórios de operação de *usinas nucleoelétricas* exigidos pela *Comissão Nacional de Energia Nuclear* (*CNEN*).

1.2 CAMPO DE APLICAÇÃO

1.2.1 Os requisitos estabelecidos nesta Norma aplicam-se às *usinas nucleoelétricas* com *Autorização para Operação Inicial ou Permanente*.

2. GENERALIDADES

2.1 INTERPRETAÇÕES

2.1.1 Qualquer dúvida, que possa surgir com referência às disposições desta Norma, será dirimida pela Comissão Nacional de Energia Nuclear (*CNEN*)

2.1.2 A *CNEN* pode, por meio de Resolução, modificar, eliminar ou acrescentar requisitos aos constantes nesta Norma, conforme considerar apropriado ou necessário.

2.2 COMUNICAÇÕES

2.2.1 Os relatórios decorrentes de disposições desta Norma devem ser endereçados à *CNEN*.

2.3 RESPONSABILIDADES

2.3.1 A *organização operadora* é a responsável pela implementação dos requisitos estabelecidos nesta Norma.

2.4 NORMAS COMPLEMENTARES

2.4.1 Onde aplicável, devem ser observados os requisitos das seguintes normas:

a. CNEN-NE-1.22: "Programas de Meteorologia de Apoio de Usinas Nucleoelétricas";
b. CNEN-NE-1.26: "Segurança na Operação de Usinas Nucleoelétricas";
c. CNEN-NE-3.01: "Diretrizes Básicas de Radioproteção";
d. CNEN-NE-1.04: "Licenciamento de Instalações Nucleares".

3. SIGLAS E DEFINIÇÕES

Para os fins desta Norma, serão adotadas as seguintes siglas e definições:

1. ***Acidente*** - desvio inesperado e significativo das condições normais de operação de uma instalação, incluindo ocorrências previstas, *acidentes postulados* ou *acidentes severos*, tais que possam resultar em danos à propriedade ou ao meio ambiente ou ainda em exposições de trabalhadores ou de indivíduos do público acima dos limites primários de dose equivalente estabelecidos pela *CNEN*.

2. ***Acidente postulado*** – acidente considerado como de ocorrência admissível para fins de análise, visando ao estabelecimento de condições de *segurança* capazes de impedir e/ou minimizar eventuais conseqüências.

3. ***Acidente severo*** – acidente que excede as *bases de projeto* e que acarreta *falhas* em estruturas, sistemas e componentes, impedindo, dessa forma, a refrigeração do núcleo do reator, conforme projetada, levando a uma degradação significativa do mesmo.

4. ***Análise de segurança*** - estudo, exame e descrição do comportamento previsto da *usina* durante toda sua vida, em situações normais, transitórias ou de *acidentes postulados*, com o objetivo de determinar:

- as margens de *segurança* previstas em *operação normal* ou em regime transitório;

- a adequação de *itens* para prevenir *acidentes* ou atenuar as conseqüências dos *acidentes* que possam ocorrer.

5. ***Autorização para operação*** - ato pelo qual a *CNEN* autoriza a operação da *usina* sob condições especificadas, podendo ser uma *Autorização para Operação Inicial* ou uma *Autorização para Operação Permanente*.

6. ***Autorização para Operação Inicial (AOI****)* – *Autorização para Operação* concedida para início da fase operacional da *usina*, após:

- verificação de que a construção está substancialmente concluída;
- completada a avaliação do *Relatório Final de Análise de Segurança – RFAS* e dos resultados dos testes pré-operacionais; e
- constatada a inclusão, na *usina*, de todas as condições suplementares de *segurança* exigidas pela *CNEN* durante a fase de construção.

7. ***Autorização para Operação Permanente (AOP****)* – *Autorização para Operação* concedida para que a *usina* possa ser operada em caráter permanente, após a conclusão da operação inicial e da operação com capacidade nominal em condições normais durante um intervalo de tempo contínuo, fixado pela *CNEN*.

8. ***Bases de projeto*** **-** conjunto de informações que identificam as funções específicas a serem desempenhadas por um *item* da *usina* e os valores específicos ou limites de variação desses valores, escolhidos para parâmetros de controle como dados fundamentais de referência para o projeto. Esses valores podem ser:

a. limitações derivadas de práticas geralmente aceitas, de acordo com o estado atual da tecnologia, para atingir objetivos funcionais; ou

b. requisitos derivados da análise (baseados em cálculos e/ou experiências) dos efeitos de *acidentes postulados* para os quais um *item* deve atingir seus objetivos funcionais.

9. ***Causa-raiz*** - É a causa fundamental de um determinado *evento* que, se eliminada, evitará a sua ocorrência ou repetição.

10. ***CNEN*** **-** Comissão Nacional de Energia Nuclear.

11. ***Combustível nuclear*** (ou simplesmente ***combustível***) - material físsil ou contendo nuclídeos físseis que**,** quando utilizado em um *reator nuclear*, possibilita uma reação nuclear em cadeia.

12. ***Comissionamento*** – processo durante o qual componentes e sistemas da *usina nucleoelétrica*, tendo sido construídos e montados, são tornados operacionais, procedendo-se à verificação de sua conformidade com as características de projeto e com os critérios de desempenho e de aceitação.

13. ***Condições de acidente*** - desvios significativos dos estados operacionais e que possam conduzir à liberação de quantidades inaceitáveis de materiais radioativos e/ou emissão de radiação, se os *dispositivos técnicos de segurança* pertinentes não funcionarem como projetados.

14. ***Condições limites para operação*** - níveis mínimos de desempenho ou de capacidade de funcionamento de sistemas ou componentes exigidos para operação segura da *usina*, conforme definidos nas *Especificações Técnicas.*

15. ***Dispositivo técnico de segurança*** - componentes, equipamentos e sistemas de *segurança* da *usina* cujo objetivo é impedir a ocorrência de *acidentes postulados* ou atenuar suas conseqüências.

16. ***Dose equivalente efetiva*** (ou simplesmente ***dose***) - grandeza expressa por:

onde:

fator de ponderação para o tecido ou órgão T, conforme estabelecido na norma CNEN-NE-3.01; e

dose equivalente média no tecido ou órgão T.

17. ***Especificações técnicas*** - conjunto de regras, aprovado pela *CNEN* no ato da *Autorização para Operação Inicial,* que estabelece limites para parâmetros, para capacidade funcional e para níveis de desempenho de equipamentos e requisitos de pessoal, visando a operação segura de *usinas nucleoelétricas.*

18. ***Evento*** operacional (ou simplesmente ***evento***) - uma *falha* ou uma seqüência de *falhas* relacionadas entre si.

19. ***Falha*** - quando um ou mais elementos básicos envolvidos na operação (equipamento, procedimento ou pessoal) não executa as funções para eles previstas.

20. ***Função de segurança*** - funções destinadas a impedir ou mitigar conseqüências de *acidentes postulados*.

21. ***Grupo critico*** - grupo de membros da população cuja exposição é razoavelmente homogênea para uma determinada fonte, sendo típica dos indivíduos que recebem as maiores *doses*, provenientes dessa fonte e relativas a um determinado caminho de exposição.

22. ***Incidente*** – *evento* sem dano significante à *usina* e/ou aos trabalhadores e ao público, mas significante em relação à *segurança* da *usina*. Pode incluir desligamentos não planejados, *paralisações forçadas e* violação de *condições limites para operação*

23. ***Item*** - termo geral que abrange qualquer estrutura, sistema, componente, equipamento, peça ou material da *usina*.

24. ***Item importante à segurança*** - *item* que inclui ou está incluído em:

    a) estruturas, sistemas e componentes cuja *falha* ou mau funcionamento pode resultar em exposições indevidas à radiação para trabalhadores ou membros do público em geral; ou

    b) estruturas, sistemas e componentes que evitam que *ocorrências operacionais previstas* resultem em *condições de acidente*; ou

    c) dispositivos ou características necessárias para atenuar as conseqüências de *falha* ou mau funcionamento de estruturas, sistemas e componentes.

25. ***Limites de segurança*** - limites impostos às variáveis operacionais importantes, considerados necessários para garantir a integridade das barreiras físicas, que protegem contra liberação não controlada de material radioativo, conforme definidos nas *especificações técnicas*.

26. ***Ocorrências operacionais previstas*** - desvios dos processos operacionais em relação à *operação normal*, que são previstos ocorrer durante a vida útil da *usina* e que, em decorrência de medidas apropriadas de projeto, não causem danos significativos a *itens importantes à segurança*, nem conduzam a *condições de acidente*.

27. ***Operação normal*** - (ou **condição normal de operação**) - operação que inclui todas as condições e *eventos* que são previstos ocorrer no curso da operação pretendida, quando realizada sob controles administrativos e de acordo com procedimentos especificados, dentro das *condições limites para operação.*

28. ***Operável*** – um sistema, subsistema, trem, componente ou dispositivo é considerado *operável* quando for capaz de executar suas funções específicas de *segurança*. Nesse caso, todos os instrumentos e controles, assim como o suprimento elétrico normal e de emergência, a água de resfriamento e de selagem e a lubrificação devem igualmente executar suas *funções de segurança* e de suporte.

29. ***Organização operadora*** - pessoa jurídica possuidora de *autorização para operação.*

30. ***Parada*** **–** desligamento programado da *usina* para troca de *combustível nuclear*.

31. ***Paralisação forçada*** **-** retirada de serviço da *usina* após a ocorrência de um *evento*, retirada essa necessária para adoção de ação corretiva conseqüente desse *evento.* Entende-se por *paralisação forçada* não apenas aquela que ocorre imediatamente após, mas também aquela que ocorre até o fim da semana seguinte à ocorrência do *evento*.

32. ***Problemas de segurança não avaliados*** **-** problemas decorrentes de modificações, ensaios, testes ou experiências que:

    a) possam aumentar a probabilidade de ocorrência ou as conseqüências de *acidentes* ou o mau funcionamento de *itens importantes à segurança* avaliados no *RFAS*;

    b) possam criar a possibilidade de *acidentes* ou de mau funcionamento, diferentes dos avaliados no *RFAS*;

    c) reduzam a margem de *segurança* definida nas *bases de projeto*, para qualquer *especificação técnica*.

33. ***Programa de Monitoração Ambiental Radiológico*** operacional ***(PMARO****)* - descrição da programação anual a ser desenvolvida, visando a monitoração do impacto ambiental causado pela operação das *usinas* existentes no local.

34. ***Programa de testes de partida*** - conjunto de testes e atividades abrangendo desde o primeiro carregamento ou recarregamento do núcleo do *reator* até a *usina* alcançar a potência nominal.

35. ***Reator nuclear*** (ou simplesmente ***reator***) **-** instalação contendo *combustível nuclear* no qual possa ocorrer processo auto-sustentado e controlado de fissão nuclear.

36. ***Redução forçada de potência*** **-** redução de potência elétrica da *usina* após a ocorrência de um *evento*, redução essa necessária para adoção de ação corretiva conseqüente desse *evento.* Entende-se por *redução forçada de potência* não apenas aquela que ocorre imediatamente mas também aquela que ocorre até o fim da semana seguinte à ocorrência do *evento*. Esta definição não inclui as reduções de potência necessárias para manutenção preventiva de rotina e atividades de calibração.

37. ***Relatório Anual de Operação*** *(****RAO****)* - relatório rotineiro, abrangendo a operação da *usina* durante o ano civil recém-findo.

38. ***Relatório Final de Análise de Segurança** (**RFAS**)* - relatório que deve ser submetido à *CNEN* a cada pedido de *Autorização para Operação Inicial*, conforme previsto na norma CNEN-NE-1.04.

39. ***Relatório de Parada** (**RP**)* – relatório emitido após cada *parada* da usina, abrangendo as atividades previstas e aquelas desenvolvidas durante esse período, assim como sua preparação e a experiência adquirida.

40. ***Relatório Mensal de Operação** (**RMO**)* - relatório rotineiro de operação, abrangendo a operação da *usina* durante o mês civil recém-findo.

41. ***Relatório de Operações Iniciais** (**ROI**)* - relatório resumido sobre o *programa de testes de partida*, incluindo testes de criticalidade, operação a baixa potência e elevação de potência até atingir o valor nominal.

42. ***Relatório do Programa de Monitoração Ambiental Radiológico*** operacional – apresenta os resultados dos parâmetros indicados no *PMARO*, no período a que se relaciona, por meio de amostras ambientais e com base nos caminhos críticos de transferência de radionuclídeos para o ser humano.

43. ***Relatório do Projeto Nuclear e Termohidráulico** (**RPNT**)* - relatório que contém o projeto nuclear e termohidráulico, com sua respectiva *análise de segurança*, correspondente a cada configuração do núcleo do *reator*.

44. ***Relatório Semestral de Rejeitos e de Liberação de Efluentes** (**RRE**)* - relatório rotineiro de operação relativo ao semestre recém-findo e abrangendo dados da *usina* sobre: geração, expedição e estocagem de rejeitos e de combustível irradiado; liberação de efluentes líquidos e gasosos e sob a forma de particulados; *dose equivalente efetiva* no *grupo crítico*; e meteorologia e fatores de dispersão e deposição atmosféricas.

45. ***Segurança técnica nuclear*** (ou simplesmente ***segurança***) - conjunto de medidas de caráter técnico, incluídas no projeto, na construção, no *comissionamento*, na manutenção e na operação da *usina*, visando evitar a ocorrência de *incidentes* e *acidentes* ou minimizar suas conseqüências.

46. ***Usina nucleoelétrica*** (ou simplesmente ***usina***) - instalação fixa, dotada de um único *reator nuclear*, para produção de energia elétrica.

4. RELATÓRIOS ROTINEIROS

4.1 RELATÓRIO DE OPERAÇÕES INICIAIS

4.1.1 A *organização operadora* deve submeter à *CNEN um Relatório de Operações Iniciais* (*ROI*) subseqüentemente:

a. à obtenção de uma *Autorização para a Operação Inicial;*
b. à modificação de *Autorização para a Operação* envolvendo aumento no nível de potência da *usina*;
c. à utilização de elemento *combustível* com projeto diferente ou fabricado por novo fornecedor; e
d. a modificações passíveis de alterar, de forma significante, o desempenho nuclear, térmico ou hidráulico da *usina*.

4.1.1.1 O *ROI* deve abordar, individualmente, os testes identificados no *Relatório Final de Análise de Segurança - RFAS* e incluir uma relação dos valores das características ou condições de operação, medidos durante o programa de testes, comparando esses valores com as previsões e especificações de projeto da *usina*.

4.1.1.2 O *ROI* deve conter a descrição de eventuais ações corretivas necessárias à operação segura da *usina*.

4.1.2 A *organização operadora* deve submeter os *ROI* à *CNEN* dentro dos seguintes prazos, prevalecendo aquele que expirar primeiro:

a. 90 (noventa) dias após a execução completa do *programa de testes de partida*; ou
b. 90 (noventa) dias após o início ou o reinicio de operação com *Autorização para Operação* Permanente; ou
c. 9 (nove) meses após a criticalidade inicial.

4.1.2.1 Se o primeiro *ROI* submetido não abranger a criticalidade inicial, a execução completa do programa de testes e o início (ou reinicio) de operação com *AOP*, devem ser apresentados relatórios suplementares, no mínimo, a cada 3 (três) meses até que todas essas etapas tenham sido incluídas.

4.2 RELATÓRIO DO PROJETO NUCLEAR E TERMOHIDRÁULICO

4.2.1 A *organização operadora* deve submeter à *CNEN* o *Relatório do Projeto Nuclear e Termohidráulico (RPNT)* a cada pedido de autorização para carregamento inicial ou recarregamento do núcleo do *reator*, incluindo eventuais revisões do *RPNT*, inclusive durante o ciclo de queima.

4.2.2 O *RPNT* deverá conter um mínimo de informações que demonstrem que os *limites de segurança,* estabelecidos nas *especificações técnicas*, não serão violados durante o ciclo de queima do *reator*.

4.2.3 O *RPNT* deverá apresentar, a cada recarga, os limites de operação do núcleo, abaixo descritos:

a. máxima densidade de potência linear;
b. fator total de canal quente;
c. fator de canal quente de entalpia nuclear;
d. limite de inserção de barras de controle;
e. queima máxima de varetas combustíveis;
f. margem de desligamento;
g. coeficiente de reatividade do combustível;
h. coeficiente de reatividade do moderador.

4.2.4 Os métodos analíticos usados para determinar os limites de operação do núcleo deverão ser aqueles previamente revistos e aprovados pela *CNEN*. Um sumário da metodologia utilizada, bem como suas referências, deverá ser apresentado no *RPNT*.

4.2.5 A *organização operadora* deve submeter o *RPNT* à *CNEN* com antecedência mínima de 3 (três) meses da criticalidade.

4.3 RELATÓRIO MENSAL DE OPERAÇÃO

4.3.1 A partir da primeira criticalidade do reator, a *organização operadora* deve submeter à *CNEN* um *Relatório Mensal de Operação* (*RMO*) sobre estatísticas de operação e experiência de desligamento, até o dia 15 do mês seguinte ao abrangido pelo relatório.

Os *RMO* devem apresentar, no mínimo, informações sobre o seguinte:

a. nível médio diário de potência;
b. dados de operação, incluindo:

i. potência autorizada;

ii. capacidade máxima confiável;
iii. potência elétrica líquida;
iv. potência restrita de operação e os motivos para restrição, se houver;
v. número de horas com gerador na linha;
vi. fatores de utilização, de disponibilidade e de capacidade do reator;

c. número de horas com o reator crítico ou desligado, energia elétrica e térmica geradas;

d. reduções da potência e desligamentos ocorridos na unidade; taxa de paralisação e desligamentos programados para o semestre que se segue;
e. caso a unidade esteja em fase de testes de elevação de potência, indicação dos valores previstos e alcançados para criticalidade, geração inicial elétrica e operação; e
f. acontecimentos notáveis com relação à monitoração ocupacional, de efluentes e ambiental, incluindo:

i. as *doses* recebidas pelas pessoas envolvidas na operação da *usina*, empregadas do operador ou não;

ii. as liberações de radioatividade ou *doses* estimadas que atinjam mais de 10% (dez por cento) dos valores anuais permissíveis; e
iii. os níveis de radioatividade detectados no meio ambiente superiores aos níveis de registro.

Nesses casos, deve-se buscar uma correlação entre a sua ocorrência e a operação da instalação.

4.4 RELATÓRIO ANUAL DE OPERAÇÃO

4.4.1 A *organização operadora* deve submeter à *CNEN* o *Relatório Anual de Operação* (*RAO*) até 1º de fevereiro do ano subseqüente àquele do relato.

4.4.1.1 O primeiro *RAO* de uma *usina* deve ser submetido até 1º de março do ano seguinte ao da criticalidade inicial.

4.4.2 O *RAO* deve fornecer um sumário abrangente da experiência operacional obtida durante o ano a que se refere, podendo envolver repetição de informações emitidas em outros relatórios, que devem ser referidos de forma clara e precisa.

4.4.2.1 Cada *RAO* deve incluir:

a. um resumo descritivo da experiência operacional, contendo também indicadores de desempenho, os quais devem ser comparados com valores internacionais .;
b. informações que descrevam, de forma completa, cada *paralisação forçada* ou *redução forçada de potência*, incluindo:

i. causas prováveis, sistemas e componentes envolvidos;

ii. ação corretiva adotada para reduzir a probabilidade de repetição do *evento;*
iii. *a* perda de geração em razão do *evento;*
iv. manutenção corretiva relacionada à *segurança*, realizada durante o *evento;*
v. identificação do caminho crítico da atividade responsável pela duração da paralisação ou da redução de potência;
vi. quaisquer exposições à radiação especificamente associadas à paralisação ocorrida, que atingirem mais de 10% (dez por cento) dos valores anuais permissíveis;

c. informações relativas às *doses equivalentes efetivas* recebidas pelas pessoas envolvidas na operação da *usina*, empregadas do operador ou não;

d. identificação de todos os *eventos* ocorridos ao longo do ano e que foram objeto de relatório, acompanhada de uma análise da experiência operacional, interna ou externa, aplicada à *usina*;
e. identificação das modificações de projeto implementadas no período;
f. indicações de *falhas* de *elementos combustíveis* resultantes de exames de *combustível* irradiado, e o método utilizado;
g. relato sobre o treinamento das equipes; envolvendo exercícios internos de emergência e de proteção contra incêndio bem como participação em exercícios gerais de emergência e em outros *eventos* considerados de importância dentro do Planejamento de Emergência, previstos para o período. Caso não tenham sido realizados, justificar;
h. resumo dos *eventos* operacionais, relacionados à *segurança*, ocorridos em outras *usinas*, no País ou no exterior, cujas análises de experiência tenham apresentado lições julgadas pela *organização operadora* como aplicáveis à *usina* objeto do relatório. Esses *eventos* e respectivas análises devem ser devidamente justificados e feitas as referências correspondentes.

4.5 RELATÓRIO SEMESTRAL DE REJEITOS E DE LIBERAÇÃO DE EFLUENTES

4.5.1 A partir da primeira criticalidade do reator, a *organização operadora* deve submeter à *CNEN* um *Relatório Semestral de Rejeitos e de Liberação de Efluentes (RRE)*, até os dias 1º de março e 1º de setembro de cada ano. Nesse relatório, devem ser especificados os efluentes líquidos, gasosos e sob forma de particulados liberados para o meio ambiente, assim como informações quanto à geração, expedição e armazenamento de rejeitos radioativos e de combustível irradiado, no período. Devem ser fornecidas, ainda, outras informações necessárias para avaliar as *doses* potenciais recebidas por indivíduos do *grupo crítico*, resultantes dessas liberações, assim como os dados meteorológicos e hidrológicos necessários ao cálculo da dispersão dos efluentes.

4.5.1.1 Os *RRE* devem apresentar, no mínimo, para cada instalação, as seguintes informações sobre liberações de efluentes:

a. medida das atividades individuais dos produtos de fissão e ativação, emissores de radiação beta e gama, liberados de modo contínuo e em bateladas, inclusive das liberações não programadas, nos efluentes líquidos;

b. medida das atividades, totais ou individuais, de emissores de radiação alfa liberados de modo contínuo e em bateladas, inclusive das liberações não programadas, nos efluentes líquidos;
c. medida das atividades individuais dos produtos de fissão e ativação, liberados, de modo contínuo e em bateladas, inclusive das liberações não programadas, nos efluentes aéreos, com as correspondentes alturas de liberação;
d. número de bateladas liberadas e tempo total de liberação dos efluentes líquidos e aéreos;
e. volume, antes da diluição, de efluentes líquidos liberados, bem como da água de diluição, correspondente ao período abrangido pelo relatório;
f. liberações não programadas de efluentes líquidos e gasosos inclusive: número de liberações e atividade liberada;
g. somatório, mês a mês, de atividades individuais de produtos de fissão e de ativação, liberadas nos efluentes líquidos, e volume de água de diluição usado durante o período;
h. somatório, mês a mês, de atividades individuais de gases nobres, particulados e trítio, liberadas nos efluentes aéreos.

4.5.1.2 Os *RRE* devem apresentar, no mínimo, para cada instalação, as seguintes informações:

a. a geração de rejeitos sólidos, incluindo o tipo de rejeito e sua composição radioquímica;
b. o número de expedições de rejeitos sólidos, a quantidade, o tipo de embalagem e sua localização;
c. listagem individual de cada embalagem contendo rejeitos radioativos, com informações relativas à composição, à atividade e ao tipo de emissão, por radionuclídeo, o nível de radiação na superfície e a um metro da mesma, o peso e número de identificação de cada embalagem;
d. o número de expedições de combustíveis irradiados.

4.5.1.3 Os *RRE* devem apresentar o cálculo da *dose equivalente efetiva* para o(s) *grupo(s) crítico(s)* e os dados necessários ao cálculo dessa *dose*, com base na metodologia aprovada pela *CNEN*.

4.5.1.4 Os *RRE* devem apresentar, no mínimo: :

a. sumário dos dados meteorológicos obtidos no período, contendo a distribuição de freqüência combinada da direção do vento por classe de velocidade do vento, para cada uma das classes de estabilidade atmosférica;
b. para o caso de liberação por batelada, os dados meteorológicos horários, citando os dias e horas de liberação. Os dados por batelada deverão ser relatados separadamente dos sumários das observações de cada trimestre;
c. para o caso de liberações não programadas, os dados meteorológicos horários, citando os dias e horas de liberação. Os dados de liberações não programadas deverão ser relatados separadamente dos sumários das observações de cada trimestre e do caso da liberação por batelada.

4.6 RELATÓRIO DO PROGRAMA DE MONITORAÇÃO AMBIENTAL RADIOLÓGICO OPERACIONAL

4.6.1 A partir da primeira criticalidade do reator, a *organização operadora* deve submeter à *CNEN* um *Relatório do Programa de Monitoração Ambiental Radiológico Operacional (PMARO)*, até o dia 1? de maio de cada ano, contendo informações necessárias ao acompanhamento do impacto radiológico ambiental causado pela operação das *usinas* existentes no local.

4.6.2 O relatório do *PMARO* deve conter, no mínimo, as seguintes informações, relacionadas ao período coberto pelo relatório:

a. efluentes liberados e rejeitos gerados neste período;
b. relação das atividades previstas no *PMARO* e cumpridas, citando as eventuais alterações feitas no programa, com justificativas;
c. resumo das principais ocorrências operacionais, que possam ter influenciado nos efluentes liberados ou na geração de rejeitos os prováveis impactos no meio ambiente provenientes dessas alterações e, quando aplicável, as medidas mitigadoras tomadas;
d. resumo das principais características do local, com ênfase nas possíveis modificações que possam ter ocorrido no período, com relação aos usos da água e do solo e aos hábitos e distribuição da população;
e. resultados das análises realizadas em cada matriz monitorada, contendo:

i. identificação do ponto de amostragem;

ii. tipo de análise realizada;
iii. unidade em que estão sendo relatados os resultados das análises, inclusive se em peso seco, peso úmido ou peso de cinzas, quando aplicável;
iv. radionuclídeos detectados com os respectivos valores de atividade medidos e erros associados;

f. resultados das monitorações feitas por meio de dosímetros termoluminescentes, ou outros tipos de detectores, indicando localização, tempo de residência, radionuclídeos detectados e sua atividade. Devem ser justificados os valores encontrados acima daqueles previstos;

g. no caso de amostras compostas, descrição do procedimento de composição das mesmas, incluindo datas ou períodos e volumes amostrados;
h. descrição dos testes estatísticos utilizados no tratamento dos dados;
i. análise comparativa entre os resultados obtidos no período, os de períodos anteriores e os da média do pré-operacional para cada matriz e, sempre que possível, para cada ponto de amostragem;
j. avaliação do impacto causado sobre o meio ambiente, correlacionando-o, sempre que possível, com ocorrências operacionais no período;
k. apresentação do desempenho obtido em Programas de Intercomparação ou equivalentes.
l. os resultados do Programa de Monitoração Ambiental do Depósito de Rejeitos Radioativos, incluindo o mapa com a situação dos locais monitorados e, para cada ponto de medida: os valores obtidos com dosímetros termoluminescentes e com monitores portáteis; os parâmetros estatísticos – número de dados, menores e maiores valores, médias, medianas e desvios padrões; e gráficos com a distribuição anual e mensal de doses

4.7. RELATÓRIO DE PARADA

4.7.1 Após cada *Parada* da usina, deverá ser apresentado um *Relatório de Parada - RP*, abrangendo:

a. o Programa ALARA, incluindo, para as atividades de maior impacto radiológico, as metas previstas e os valores atingidos para a dose coletiva e a dose média dos trabalhadores. Em caso de não alcance dessas metas, esse fato deve ser justificado;

b. as principais ocorrências com impacto na área de Proteção Radiológica;
c. a quantidade de rejeitos radioativos sólidos gerados, por tipo de rejeito;
d. as atividades dos efluentes líquidos e gasosos liberados e a dose *no grupo crítico*;
e. as modificações de projeto implementadas;
f. o programa de treinamento; e
g. os resultados da Inspeção em Serviço.

4.7.2 Referir como foram utilizadas as experiências de *Paradas* anteriores na atual programação.

4.7.3 Apresentar uma visão crítica das atividades desenvolvidas, incluindo seus pontos positivos e negativos, e as recomendações para a próxima *Parada*, objetivando a redução de doses;

4.7.4 Comparar os valores atingidos durante a *Parada* com aqueles obtidos em *Paradas* anteriores, em termos de dose coletiva. .

5. NOTIFICAÇÃO DE EVENTOS

5.1 REQUISITOS GERAIS

1. A *organização operadora* deverá notificar à *CNEN* a ocorrência de:

a. declaração de situação de emergência;

b. *eventos* não iniciadores de situação de emergência ou que ainda não tenham sido identificados como tal.

2. A *organização operadora* deverá manter um canal de comunicação aberto e contínuo com a *CNEN*, durante o curso do *evento* ou da situação de emergência.

3. A *organização operadora* deverá notificar à *CNEN*, em tempo hábil, durante o curso do *evento* ou da situação de emergência:

a. qualquer degradação adicional no nível de *segurança* ou nas condições da *usina*, incluindo aquelas que requeiram ativação do Plano de Emergência, caso esse não tenha sido ativado anteriormente;

b. qualquer mudança de uma classe de emergência para outra;
c. informações sobre parâmetros relevantes para a caracterização do nível de *segurança* da *usina*;
d. resultados relevantes de análises de condições e comportamento da *usina*;
e. a identificação da necessidade da adoção de qualquer ação de proteção, que divirja de uma condição da *Autorização para Operação* ou de uma *especificação técnica* necessária para proteger a saúde e a *segurança* do público. Essa notificação deve ser feita antes da ação protetora ser executada ou, não sendo exeqüível, o mais breve possível após o fato consumado;
f. a eficácia das respostas automáticas ou das medidas de proteção tomadas;
g. informações relativas ao comportamento da *usina* que não tenha sido adequadamente interpretado;
h. a declaração de encerramento do *evento* ou da situação de emergência.

5.2 EVENTOS INICIADORES DE SITUAÇÃO DE EMERGÊNCIA

A *organização operadora* deve notificar à *CNEN*, tão logo quanto possível e no prazo máximo de uma hora, a declaração de qualquer situação de emergência.

5.3 EVENTOS NÃO INICIADORES DE SITUAÇÃO DE EMERGÊNCIA

Quando o *evento* não promover a declaração de situação de emergência ou ainda não tenha sido identificado como tal, a O*rganização Operadora* deve notificar à *CNEN*, tão logo quanto possível e no prazo máximo de uma ou quatro horas, conforme aplicável, a ocorrência de qualquer dos *eventos* relacionados abaixo:

5.3.1 ***Notificações em até 1 hora:***

1. Início de qualquer desligamento da unidade requerido pelas *especificações técnicas.*

2. Qualquer desvio das e*specificações técnicas,* realizado intencionalmente, necessário para garantir a *segurança* do público e do meio ambiente, quando nenhuma ação, consistente com as condições da *Autorização para Operação*, puder proporcionar, de imediato, uma proteção adequada ou equivalente.

3. Qualquer *evento* ou condição, durante a operação, que possa levar a *usina*, incluindo suas principais barreiras de *segurança*, a uma condição seriamente degradada, incluindo:

a. uma condição não analisada que comprometa significativamente a sua *segurança*; ou

b. uma condição fora das *bases de projeto*; ou

c. uma condição não coberta por procedimentos de operação ou de emergência.

4. Qualquer fenômeno natural ou outra condição externa, que possa levar a uma ameaça real à *segurança* da *usina*, ou dificulte significativamente a execução das tarefas necessárias à operação segura da mesma.

5. Circunstância de perda ou redução da capacidade de avaliação de uma eventual emergência.[1]

6. Circunstância de perda ou redução da capacidade de comunicação ou acesso, que possa comprometer a execução do Plano de Emergência Local.

7. Qualquer *evento* que possa levar a uma ameaça real à *segurança* da *usina* ou dificulte significativamente a execução das atividades necessárias para a operação segura da *usina*, incluindo incêndio, liberação de gases tóxicos e ou liberação de radioatividade.

5.3.2 ***<u>Notificações em até 4 horas:</u>***

1. Qualquer *evento* em que, estando o reator desligado, tenha sido identificada uma configuração que, caso o reator estivesse em operação, teria colocado a *usina*, incluindo suas principais barreiras de *segurança*, numa condição seriamente degradada ou em uma condição não analisada que poderia comprometer significativamente a *segurança* da *usina*.

2. Qualquer *evento* ou condição que resulte em uma atuação automática ou manual indevida, de qualquer *função de segurança*, incluindo o sistema de proteção do reator, exceto quando a atuação resulta de uma atividade previamente planejada durante testes ou operação do reator.

3. Qualquer *evento* ou condição que, isoladamente, poderia ter impedido o cumprimento das *funções de segurança* das estruturas ou sistemas necessários para:

a. desligar o reator e mantê-lo em uma condição segura em desligado; ou

b. emover o calor residual; ou

c. controlar a liberação de material radioativo; ou
d. mitigar as conseqüências de um *acidente*.

6. RELATÓRIOS DE EVENTOS

6.1 A *organização operadora* deve submeter à *CNEN* um relatório dos *eventos*, no prazo máximo de 30 dias após a descoberta do mesmo, para qualquer *evento* classificado segundo a subseção 6.2. desta norma.

6.1.1 A *organização operadora* deve relatar o *evento* independentemente do modo de operação da *usina,* do nível de potência, da estrutura, do sistema ou do componente onde se iniciou o *evento*.

6.1.2 A *organização operadora* deve submeter à *CNEN*, no prazo de 30 dias, um relatório dos *eventos* que não se enquadrem nas classes da subseção 6.2, sempre que o *evento* seja considerado relevante para a *segurança*.

6.2. CLASSIFICAÇÃO DE EVENTOS

6.2.1 A *organização operadora* deve classificar os *eventos* da seguinte forma:

a) **Classe 1**: *Eventos* relacionados às e*specificações técnicas*:

qualquer desligamento requerido pelas *especificações técnicas*; ou

qualquer operação ou condição não permitida pelas e*specificações técnicas*, de acordo com a norma CNEN-NE-1.26, item 4.6; ou

qualquer desvio das e*specificações técnicas,* realizado intencionalmente, necessário para garantir a *segurança* do público e do meio ambiente, quando nenhuma ação, consistente com as condições da *Autorização para Operação*, puder proporcionar de imediato uma proteção adequada ou equivalente, conforme a norma CNEN-NE-1.04.

b) **Classe 2**: Qualquer *evento* que tenha levado a *usina*, incluindo suas principais barreiras de *segurança,* a uma condição seriamente degradada ou a:

1) uma condição não analisada que tenha comprometido de forma significativa a *segurança* da *usina*;

2) uma condição fora das *bases de projeto* da *usina*; ou

3) uma condição não coberta pelos procedimentos de operação ou de emergência da *usina.*

c) **Classe 3**: Qualquer fenômeno natural ou outra condição externa que tenha levado a uma ameaça real à *segurança* da *usina,* ou dificultado a execução das tarefas necessárias à operação segura da *usina*.

d) **Classe 4**: Qualquer *evento* que tenha resultado em uma atuação automática ou manual de qualquer *dispositivo técnico de segurança*, incluindo o sistema de proteção do *reator*, exceto quando:

1) a atuação resultou de uma atividade previamente planejada durante testes ou operação do *reator*;

2) a atuação não foi prevista e:

i. ocorreu enquanto o sistema estava corretamente fora de serviço; ou
ii. ocorreu após a *função de segurança* ter sido realizada.

e) **Classe 5**: Qualquer *evento* que, sozinho, poderia ter impedido a realização da *função de segurança* de sistemas e estruturas, necessária para:

1. desligar o *reator* e mantê-lo numa condição segura em desligado; ou
2. remover o calor residual; ou
3. controlar a liberação de material radioativo; ou
4. mitigar as conseqüências de um *acidente*.

(2)

f) **Classe 6**:: Qualquer *evento* no qual uma causa única ou condição que tenha provocado a inoperabilidade de um componente ou grupo de componentes em um único trem ou canal redundante em mais de um sistema de *segurança*, ou a inoperabilidade de canais ou trens redundantes em um mesmo sistema.

1) desligar o *reator* e mantê-lo numa condição segura em desligado; ou

2) remover o calor residual; ou

3) controlar a liberação de material radioativo; ou

4) mitigar as conseqüências de um *acidente*.

g. **Classe 7**: Qualquer liberação de efluentes radioativos por via aérea, que possa expor o público a níveis de radiação que acarretam doses superiores aos limites primários estabelecidos para indivíduos do público, de acordo com a Norma CNEN-NE-3.01.

h. **Classe 8**: Qualquer liberação de efluentes líquidos radioativos, que possa expor o público a níveis de radiação que acarretem doses superiores aos limites primários estabelecidos para indivíduos do público, de acordo com a Norma CNEN-NE-3.01.

i. **Classe 9**: Qualquer *evento* que tenha levado a uma ameaça real à *segurança* da *usina* ou que tenha dificultado significativamente os seus trabalhadores de executarem as tarefas necessárias para a operação segura do *reator*, incluindo incêndio, liberação de gases tóxicos ou liberação de radioatividade.

6.3 REQUISITOS DE RELATÓRIO

6.3.1 IDENTIFICAÇÃO

O relatório deverá conter os seguintes requisitos de identificação:

a. Nome da unidade onde o *evento* ocorreu;
b. Título do *evento*, incluindo uma descrição concisa do principal problema ou assunto associado ao *evento*;
c. Data do *evento*;
d. Número do relatório;
e. Modo de operação da unidade - como definido nas *Especificações Técnicas* - no momento em que ocorreu o *evento*;
f. Percentual da potência nuclear autorizada na qual o reator estava operando quando ocorreu o *evento*;
g. Classificação do *evento* segundo a escala INES da Agência Internacional de Energia Atômica;
h. Quando o *evento* for classificado como emergência, colocar a identificação da classe de acordo com o plano de emergência;
i. Classe de *evento* desta Norma em que o mesmo se enquadra e que requereu a emissão do relatório;

6.3.2 CONTEÚDO

O conteúdo do relatório deve incluir:

a. Uma descrição do *evento,* contendo:

1. Uma narração clara e específica do *evento* tal que os leitores familiarizados com o projeto de *reatores nucleares*, mas não familiarizados com o projeto da *usina* em particular, possam entendê-lo completamente . Essa descrição, sob o ponto de vista da operadora, deve incluir desenhos, figuras, gráficos, tabelas, fotografias e outros recursos que permitam um completo entendimento do *evento*.

2. As seguintes informações específicas sobre o *evento* em questão:

i. condições de operação da *usina* antes do *evento*;

ii. condições das estruturas, componentes ou sistemas que estavam inoperáveis no início do *evento* e que contribuíram para o mesmo;
iii. data e hora aproximada das ocorrências;
iv. a *causa-raiz* de cada *falha* de componente ou sistema ou de erro pessoal, se conhecida;
v. o modo de *falha*, o mecanismo (causa imediata) e/ou o efeito de cada componente que falhou, se conhecidos;
vi. a função de cada componente e o nome dos sistema referidos no relatório, de acordo com a nomenclatura utilizada na *usina*.
vii. para *falhas* de componentes com múltiplas funções, a inclusão da lista dos sistemas ou funções secundárias que também foram afetados;
viii. para *falhas* que causaram a inoperabilidade de um trem de um sistema de *segurança*, o tempo estimado desde a descoberta da *falha* até o trem ter retornado à condição de *operável*;
ix. o método de descoberta de cada *falha* do componente ou sistema ou do erro de procedimento;
x. as ações do operador que afetaram o curso do *evento*, incluindo erros de operadores, deficiências em procedimentos, ou ambos, que contribuíram para o *evento*. Para cada erro de operador, a *organização operadora* deve discutir:

- se o erro foi um erro cognitivo (por exemplo, *falha* em reconhecer a condição atual da *usina* ou a natureza do evento *ou* em perceber quais sistemas deveriam estar funcionando,) ou um erro de procedimento;

- se o erro foi contrário ao estabelecido em um procedimento aprovado, se foi um resultado direto de um erro em um procedimento aprovado ou se estava associado com uma atividade ou tarefa não coberta por um procedimento aprovado;
- qualquer característica não usual do local de trabalho (por exemplo, calor, ruído) que diretamente contribuiu para o erro; e
- a qualificação do pessoal envolvido;

xi. respostas de sistemas de *segurança* iniciadas automática ou manualmente e;

xii. a identificação de cada componente que falhou durante o *evento*;

b. Uma avaliação das conseqüências do *evento* para a *segurança* e suas implicações. Essa avaliação deve incluir a disponibilidade de outros sistemas ou componentes que poderiam ter realizado a mesma função que aqueles que falharam durante o *evento*;

c. A descrição das ações corretivas planejadas como resultado do *evento*, incluindo aquelas que objetivam reduzir a probabilidade de que *eventos* similares ocorram no futuro;
d. Referência a *eventos* similares ocorridos anteriormente na *usina*, discutindo, quando for o caso, o porquê das ações corretivas adotadas não terem evitado a repetição do *evento*.

6.4 DISPOSIÇÕES COMPLEMENTARES

A *organização operadora* poderá requerer à *CNEN*, com uma justificativa adequada, exceções aos requisitos de relatório.

COMISSÃO DE ESTUDO

Presidente:

Wilson Melo da Silva Filho SENOR

Membros:

Paulo Cabrera Ambros SLC/CNEN

Iara Monteiro SLC/CNEN

José Gonçalves Dias PJU/CNEN

Maisa Beraldino PJU/CNEN

Vaner O. Souza Eletronuclear

José Arnaldo M. de C. Ochotorena Eletronuclear

Jorge Eduardo de Souza Mendes Eletronuclear

Hamilton de Lima Rodrigues dos Santos Eletronuclear

Vaner O. Souza Eletronuclear

ANEXO III

Brazilian Nuclear Power Utility Chooses Verity for its Intranet Web Site[107]

ELETRONUCLEAR Selects Verity K2 Enterprise Solution for its Search Relevance, and Categorization and Scalability Capabilities

SUNNYVALE, Calif. - 2005-09-26 - Verity Inc. (NASDAQ: VRTY), a leading provider of enterprise search software that enables organizations to discover, analyze and process all the digital information within their enterprises, today announced that Eletrobrás Termonuclear S.A., a nuclear power utility whose majority owner is the Brazilian government, has implemented Verity® K2 Enterprise (K2E) software to provide search and classification on its main intranet Web site that connects to different data sources, including its operating reports.

The utility, known as ELETRONUCLEAR, sought an enterprise search solution that would help to improve its employees' productivity, while extending its world-class safety standards to the company's intranet Web site.

Nuclear utilities often draw on their own operating experience as well as the experience of similar companies around the globe. ELETRONUCLEAR combines its own operating experience with the collective experience of the industry through the World Association of Nuclear Operators in its knowledge repositories to allow its personnel to draw on a vast amount of intellectual capital with speed and accuracy.

"Brazil's ELETRONUCLEAR relies on just-in-time information management, providing our employees with the most up-to-date information 24x7. Efficient and fast delivery of relevant information is the key benefit of the Verity technology" said Luiz Celso Oliveira, ELETRONUCLEAR's external operating experience coordinator. "The utility's operating reports are expected to be made readily available, and used by a broad spectrum of our personnel to support safe and reliable performance of the nuclear power utility. Acquiring and using relevant operating experience is a job expectation for many positions across the utility. Verity K2E is an effective and easy-to-use enterprise search solution. We expect that it will be widely used by our personnel to enhance their productivity and job performance."

ELETRONUCLEAR evaluated other search solution options. However the advanced search and classification capabilities of the Verity software were seen as clearly superior to the others considered. Further, the scalability of the K2E software to handle content growth made it the solution that best fits ELETRONUCLEAR's business needs and requirements.

"ELETRONUCLEAR is a standard bearer in operating excellence for nuclear power utilities," said Anthony J. Bettencourt, Verity's chief executive officer. "It sought to

---

[107] Fonte: http://www.autonomy.com/content/News/Releases/2005/V0926.en.html. Acesso em 10 fev 2007

improve the quality of search on the company's internal Web presence to properly reflect its operating reputation. We feel honored that an organization that demands excellence of itself saw excellence in Verity K2 Enterprise software."

About ELETRONUCLEAR

ELETRONUCLEAR - Eletrobrás Termonuclear S.A. is a nuclear power utility, whose goal is to produce electricity with high standards of safety and efficiency, at a competitive cost. It is a mixed-economy company, owned in its majority by the Brazilian government. It is a subsidiary of Eletrobrás, the holding company of the Brazilian electrical sector.

ELETRONUCLEAR was created in 1997 as a result of a merger of the nuclear area of Furnas Centrais Elétricas with Nuclen Engenharia e Serviços. Its staff concentrates the best skills in the design, construction and operation of thermonuclear power plants, adjusting its human resources and technological capacity to the new challenges.

ELETRONUCLEAR is the owner and operator of the Almirante Álvaro Alberto Nuclear Power Station, also known simply as the Angra Nuclear Power Plant, which consists of the Angra 1 and Angra 2 plants, in operation, and Angra 3, which is pending a governmental approval to resume construction and whose environmental impact studies for licensing purposes are presently under preparation.

About Verity

With headquarters in Sunnyvale, California, Verity provides enterprise search software that enables organizations to discover, analyze and process all the information within their enterprises. The company's solutions provide integrated search, classification, recommendation, monitoring and analytics across the real-time flow of enterprise information, along with self-service Q & A. In addition, Verity's business process management and content capture solutions activate information and accelerate its flow from person to person and between systems. Verity technology also serves as a core component of more than 260 applications from leading independent software vendors.

Around the world, more than 15,000 organizations of all sizes and types rely on Verity to manage their intellectual capital. In the private sector, these include ABB, AT&T, AXA, Bristol-Myers Squibb, Capgemini, Capital One, Deloitte Consulting, EDGAR Online, Ford, GMAC, Hewlett-Packard, Kaiser Permanente, Mayo Foundation, PricewaterhouseCoopers, SAP, UBS and Verizon. In the public sector, customers come from all levels and branches of government, including the U.S. General Services Administration, the U.S. Air Force and Army, the U.S. Department of Defense, including the armed forces units, and the U.S. Departments of Energy and Justice. Independent software vendors that integrate Verity include EMC Documentum, FileNet, Lotus, Oracle, Stellent, TIBCO and Xerox DocuShare.

ANEXO IV

Extrato da Guideline da WANO (World Association of Nuclear Operators) usada como referência para atividades de gestão da Experiência Operacional[108].

Effectively using operating experience includes analysing both internal and external operating experience to identify fundamental weaknesses and then determining appropriate plant specific corrective actions that will minimise the likelihood of similar events.

The plant's goal for operating experience (OE) is to effectively and efficiently use lessons learned from plant and external operating experience to improve plant safety and reliability. Learning and applying the lessons from operating experience is an integral part of plant culture and is encouraged by managers throughout the organization.

Plant personnel regard operating experience as helpful and important to them and they use this information at every opportunity. Methods of using operating experience are structured to provide applicable information to the right personnel in time to make a difference. When plant personnel analyse the causes of significant plant events, operating experience is routinely reviewed to determine if and why previous lessons were not effectively learned.

Management expectations for distributing and using operating experience information include the following elements:

- Review and screen plant and external operating experience information in a timely manner.
- Establish criteria to help determine what external operating experience applies to the plant.
- Make applicable external operating experience widely available to plant personnel in a manner that encourages its routine use.
- Distribute applicable external operating experience to the appropriate personnel for review, analysis and use.
- Trend plant events and near misses to identify recurring issues.
- Investigate and identify causes of plant events and near misses.

[108] Fonte: WANO/ELETRONUCLEAR

Screening of Operating Experience

A. INTRODUCTION

Internal events

Plant event reports including minor events and near miss reports are screened in order to decide if an event should be further analysed or not, the necessary level and priority of analysis of the event, the need to communicate the facts of the event to plant personnel, potential adverse trends and if the event needs to be communicated externally.

External events

The primary objective of screening external operating experience is to identify and transfer lessons learned from other plants into actions that enhance each plant's safety and reliability. External experience is screened for applicability, significance and the potential that a similar event could occur at the plant. The screening process identifies issues requiring more in-depth evaluation.

B. DISCUSSION

Internal events

Guidance is provided to screen event and near miss reports and determine a priority for analysis based on risk and consequence to the plant. The screening process has welldefined thresholds so that events important only for trending (apparent cause) are sorted out from significant events that require a formal event analysis (root cause). This process promotes using resources so that significant events are thoroughly analysed. Personnel assigned to screen events have the necessary training and qualifications.